

7 Fundamental thresholds in compressed sensing: a high-dimensional geometry approach

Weyu Xu and Babak Hassibi

In this chapter, we introduce a unified high-dimensional geometric framework for analyzing the phase transition phenomenon of ℓ_1 minimization in compressive sensing. This framework connects studying the phase transitions of ℓ_1 minimization with computing the Grassmann angles in high-dimensional convex geometry. We demonstrate the broad applications of this Grassmann angle framework by giving sharp phase transitions for ℓ_1 minimization recovery robustness, weighted ℓ_1 minimization algorithms, and iterative reweighted ℓ_1 minimization algorithms.

7.1 Introduction

Compressive sensing is an area of signal processing which has attracted a lot of recent attention for its broad applications and rich mathematical background [7] [19] and Chapter 1. In compressive sensing, we would like to recover an $n \times 1$ real-numbered signal vector x , but we can only get $m < n$ measurement samples through a linear mixing of x . Namely

$$y = Ax, \tag{7.1}$$

where A is an $m \times n$ measurement matrix and y is an $m \times 1$ measurement result. In an ideal model for compressive sensing, x is an $n \times 1$ unknown k -sparse signal vector, which is defined as a vector having only k nonzero elements. This special structure of x makes recovering x from the compressed measurement y possible.

A naive way to decode or solve for the k -sparse x from y is to enumerate the $\binom{n}{k}$ possible supports of x and then try to see whether there exists such an x satisfying $y = Ax$. But this is of exponential complexity if k is proportionally growing with n and is not computationally feasible. What enables practical compressive sensing is the existence of

efficient decoding algorithms to recover the sparse signal x from the compressed measurements y . Arguably the most prominent and powerful decoding approach is the Basis Pursuit programming, namely the ℓ_1 minimization method [13, 15]. The ℓ_1 minimization method solves the following problem

$$\begin{aligned} \min & \|z\|_1 \\ \text{subject to } & y = Az, \end{aligned} \quad (7.2)$$

where $\|z\|_1$ denotes the ℓ_1 norm of z , namely the sum of the absolute values of all the elements in z . This is a convex program which is easy to solve. It has been empirically observed to work very well in producing sparse solutions. Breakthroughs, for example, [11][15][20], in understanding why ℓ_1 minimization successfully promotes sparsity have emerged in recent years, and have triggered an explosive growth of research in compressive sensing, see Chapter 1.

We should remark that in this chapter we are particularly interested in the parameter regime where k and m grow proportionally with n , as n grows large. In other words, the number of measurements is $m = \delta n$, and the number of nonzero elements of x is $k = \rho \delta n = \zeta n$, where $0 < \rho < 1$ and $0 < \delta < 1$ are constants independent of n , and $\delta > \zeta$.

It has been empirically observed and theoretically shown [15, 20] that the ℓ_1 minimization method often exhibits a “phase transition” phenomenon: when the signal support size is below a certain threshold, ℓ_1 minimization will recover the signal vector with overwhelming probability; while when the signal support size is above this threshold, ℓ_1 minimization will fail to recover the signal vector with high probability. Studying this phase transition phenomenon and characterizing the threshold for the support size k has been a very important and active research branch in the development of compressive sensing theories [15, 20] [4] [29, 43, 44] [28, 40, 45, 46][21, 38, 39] [34, 36] [16, 17] and Chapter 9. This branch of research gives precise prediction of sparse recovery algorithms, brings theoretical rigor to compressive sensing theories, and inspires new powerful sparse recovery algorithms.

The first work in the literature that precisely and rigorously characterized the phase transition was [15, 20], through beautifully connecting the projection of high-dimensional convex polytopes and the success of ℓ_1 minimization. In [15, 20], Donoho and Tanner formulated a k -neighborly polytope condition on the measurement matrix A for ℓ_1 minimization to generate the original sparse signal. As shown in [15], this k -neighborly polytope A is in fact a necessary and sufficient condition for (7.2) to produce the sparse solution x satisfying (7.1). This geometric insight, together with known results on the neighborliness of projected polytopes in the literature of convex geometry [1, 41], has led to sharp bounds on the performance of ℓ_1 minimization. In [15], it was shown that if the matrix A has i.i.d. zero-mean Gaussian entries, then the k -neighborly polytope condition holds with overwhelming probability if k is sufficiently small. In the linear scaling setting for m , n , and k discussed in this chapter, the relation between m , n , and k in order for the k -neighborly polytope condition to hold is precisely characterized and calculated in [15]. In fact, the computed values of ζ for the so-called “weak”

threshold, obtained for different values of δ through the neighborly polytope condition in [15], match exactly with the phase transitions obtained by simulation when n is large.

However, the neighborly polytope approach in [15] only addressed the phase transitions for ideally sparse signal vectors whose residual elements are exactly zero excluding the k nonzero components. By comparison, the popular restricted isometry property (RIP) [11][4] and Chapter 1 can also be used to analyze the robustness of ℓ_1 minimization [10], even though the RIP analysis generally produces much looser phase transition results than the neighborly polytope condition [4]. Then the question is whether we can have a unified method of determining precise phase transitions for ℓ_1 minimization in broader applications. More specifically, this method should give us tighter phase transitions for ℓ_1 minimization than the RIP condition; but it should also work in deriving phase transitions in more general settings such as:

- phase transitions for recovering approximately sparse signals, instead of only perfectly sparse ones [43];
- phase transitions when the compressed observations are corrupted with noises [42];
- phase transitions for weighted ℓ_1 minimization, instead of regular ℓ_1 minimization [29];
- phase transitions for iterative reweighted ℓ_1 algorithms [12][44].

In this chapter, we are interested in presenting a unified high-dimensional geometric framework to analyze the phase transition phenomenon of ℓ_1 minimization. As we will see, in many applications, it turns out that the performance of ℓ_1 minimization and its variants often depends on the null space “balancedness” properties of the measurement matrix A , see Chapter 1. This unified high-dimensional geometric analysis framework investigates the phase transitions for the null space “balancedness” conditions using the notion of a Grassmann angle. This framework generalizes the neighborly polytope approach in [15, 20] for deriving phase transitions of recovering perfectly sparse signals; however, this Grassmann angle framework can be further used in analyzing the performance thresholds of ℓ_1 minimization for *approximately sparse* signals, weighted ℓ_1 minimization algorithms, and iterative reweighted ℓ_1 minimization algorithms. In this chapter, we will present the Grassmann angle framework for analyzing the null space “balancedness” properties in detail by focusing on the example of characterizing the threshold bounds for ℓ_1 minimization robustness in recovering approximately sparse signals. Then we will briefly illustrate the application of this Grassmann angle framework in characterizing the phase transitions for weighted ℓ_1 minimization and iterative reweighted ℓ_1 minimization algorithms. This framework and results of this chapter have earlier appeared in [43, 29, 44].

Before demonstrating how the Grassmann angle geometric framework can be used to analyze the null space “balancedness” properties, we will give an overview of the main results, comparisons of this Grassmann angle approach with other approaches in the literature, the geometrical concepts to be used frequently in this chapter, and also the organization of this chapter.

7.1.1 Threshold bounds for ℓ_1 minimization robustness

Instead of assuming that x is an exactly k -sparse signal, we now assume that k components of x have large magnitudes and that the vector comprised of the remaining $(n - k)$ components has an ℓ_1 -norm less than some value, say, $\sigma_k(x)_1$. We will refer to this type of signal as an approximately k -sparse signal, or for brevity only an approximately sparse signal. It is also possible that the y can be further corrupted with measurement noise. In this case exact recovery of the unknown vector x from a reduced number of measurements is generally not possible. Instead, we focus on obtaining a reconstruction of the signal that is “close” to the true one. More precisely, if we denote the unknown signal as x and denote \hat{x} as one solution to (7.2), we prove that for any given constant $0 < \delta < 1$ and any given constant $C > 1$ (representing how close in ℓ_1 norm the recovered vector \hat{x} should be to x), there exists a constant $\zeta > 0$ and a sequence of measurement matrices $A \in \mathbb{R}^{m \times n}$ as $n \rightarrow \infty$ such that

$$\|\hat{x} - x\|_1 \leq \frac{2(C+1)\sigma_k(x)_1}{C-1}, \quad (7.3)$$

holds for all $x \in \mathbb{R}^n$, where $\sigma_k(x)_1$ is the minimum possible ℓ_1 norm value for any $(n - k)$ elements of x (recall $k = \zeta n$). Here ζ will be a function of C and δ , but independent of the problem dimension n . In particular, we have the following theorem.

THEOREM 7.1 *Let n, m, k, x, \hat{x} and $\sigma_k(x)_1$ be defined as above. Let K denote a subset of $\{1, 2, \dots, n\}$ such that $|K| = k$, where $|K|$ is the cardinality of K , and let K_i denote the i th element of K and $\bar{K} = \{1, 2, \dots, n\} \setminus K$.*

Then the solution \hat{x} produced by (7.2) will satisfy

$$\|\hat{x} - x\|_1 \leq \frac{2(C+1)\sigma_k(x)_1}{C-1}. \quad (7.4)$$

for all $x \in \mathbb{R}^n$, if for all vectors $w \in \mathbb{R}^n$ in the null space of A , and for all K such that $|K| = k$, we have

$$C\|w_K\|_1 \leq \|w_{\bar{K}}\|_1, \quad (7.5)$$

where w_K denotes the part of w over the subset K .

Furthermore, if $A \in \mathbb{R}^{m \times n}$ a random matrix with i.i.d. standard Gaussian $\mathcal{N}(0, 1)$ entries, then as $n \rightarrow \infty$, for any constant $C > 1$ and any $\delta = m/n > 0$, there exists a $\zeta(\delta, C) = k/n > 0$, so that both (7.4) and (7.5) hold with overwhelming probability.

As we said, the generalized Grassmann angle geometric framework can be used to analyze such null space “balancedness” conditions in (7.5), thus establishing a sharp relationship between δ, ζ , and C . For example, when $\delta = m/n$ varies, we have Figure 7.1 showing the tradeoff between the signal sparsity ζ and the parameter C , which determines the robustness¹ of ℓ_1 minimization. We remark that the above theorem clearly subsumes

¹ The “robustness” concept in this sense is often called “stability” in other papers, for example, [7].

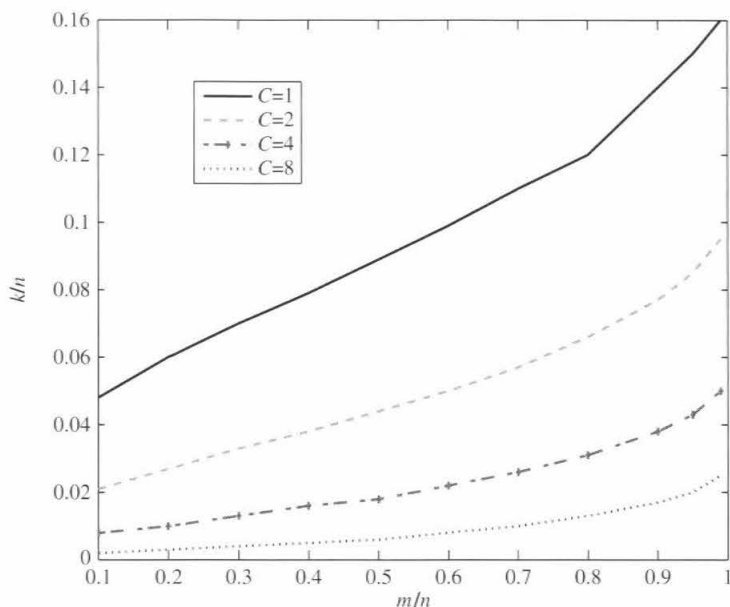


Figure 7.1 Tradeoff between signal sparsity and ℓ_1 recovery robustness as a function of C (allowable imperfection of the recovered signal is $\frac{2(C+1)\sigma_k(x)_1}{C-1}$).

perfect recovery in the perfectly k -sparse setting. For the perfectly k -sparse signal, $\sigma_k(x)_1 = 0$, and so from (7.4) we have $\|\hat{x} - x\|_1 = 0$ and therefore $\hat{x} = x$, when we allow $C \rightarrow 1$. This null space characterization is in fact equivalent to the neighborly polytope characterization from [15] in the perfectly k -sparse case when $C = 1$.

The Grassmann angle framework can be applied to characterize sparse recovery performance when there are observation noises involved and $y = Ax + e$, where e is an $m \times 1$ additive noise vector with $\|e\|_2 \leq \epsilon$. We can still use (7.2) to solve for \hat{x} . As long as A has rank m with its smallest nonzero singular value $\sigma_m > 0$, there always exists an $n \times 1$ vector Δx such that $\|\Delta x\|_2 \leq \frac{1}{\sigma_r} \epsilon$ and $y = A(x + \Delta x)$. Namely y can be seen as generated by a perturbation of x . This added perturbation Δx can then be treated in the same way as x_K in analyzing $\|x - \hat{x}\|_1$. To further bound $\|x - \hat{x}\|_2$, one can use the almost Euclidean property described in Subsection 7.1.3. For more details on dealing with observation noises based on the Grassmann angle analysis, please refer to [42].

Besides this concept of robustness in ℓ_1 norm discussed in Theorem 7.2, there are also discussions of robustness in general ℓ_p norms, which will involve other types of null space properties. The readers are encouraged to refer to Chapter 1 for an overview of this topic. A similar formulation of ℓ_1 norm robustness recovering an approximately sparse problem was considered in [14], where the null space characterization for recovering approximately sparse signals was analyzed using the RIP condition; however, no explicit values of ζ were given. Through message passing analysis, [16] recently

deals with a related but different problem formulation of characterizing the trade-off between signal sparsity and noise sensitivity of the LASSO recovery method. An overview of this new message passing analysis approach is presented in Chapter 9 of this book.

7.1.2 Weighted and iterative reweighted ℓ_1 minimization thresholds

When there is statistical prior information about the signal x , a better decoding method would be to solve the weighted ℓ_1 minimization programming [29]. The weighted ℓ_1 minimization solves the following general version of ℓ_1 minimization

$$\min_{Az=y} \|z\|_{w,1} = \min_{Az=y} \sum_{i=1}^n w_i |z_i|, \quad (7.6)$$

where the weights w_i can be any non-negative real numbers to accommodate prior information. For example, if the prior information shows that x_i is more likely to be zero or small, then a larger corresponding weight can be applied to suppress its decoding result to be zero or a small number.

Again, the successful decoding condition for weighted ℓ_1 minimization is a weighted version of the null space “balancedness” condition. It turns out that the null space Grassmann angle framework can also be readily applied to give sharp sparse recovery threshold analysis for weighted ℓ_1 minimization algorithms [29], where better phase transitions are shown when prior information is available. When no prior information about the signal is available, this null space Grassmann angle framework can also be used to analyze iterative reweighted ℓ_1 minimization [44], where we can rigorously show for the first time that iterative reweighted ℓ_1 algorithms, compared with the plain ℓ_1 minimization algorithms, can increase the phase transition thresholds for interesting types of signals.

7.1.3 Comparisons with other threshold bounds

In this section, we will review other approaches to establishing sparse recovery threshold bounds for ℓ_1 minimization, and compare their strengths and limitations.

Restricted isometry property

In [9][11], it was shown that if the matrix A satisfies the now well-known restricted isometry property (RIP), then any unknown vector x with no more than $k = \zeta n$ nonzero elements can be recovered by solving (7.2), where ζ is an absolute constant as a function of δ , but independent of n , and explicitly bounded in [11]. Please see Chapter 1 for the definition of the RIP conditions and its applications. However, it should be noted that the RIP condition is only a sufficient condition for ℓ_1 minimization to produce a sparse solution to (7.2). Partially because of this fact, the threshold bounds on ζ obtained by the RIP condition are not very sharp so far and are often a very small fraction of the bounds on ζ obtained by the neighborly polytope approach or its generalization to the Grassmann angle approach in this chapter.

One strength of the RIP condition lies in its applicability to a large range of measurement matrices. It turns out that for measurement matrices with i.i.d. zero-mean Gaussian entries, measurement matrices with i.i.d. Bernoulli entries, or matrices for random Fourier measurements, the RIP condition holds with overwhelming probability [11, 2, 33]. In contrast, the neighborly polytope approach and the Grassmann angle approach so far only rigorously work for the measurement matrices with i.i.d. Gaussian entries, even though the universality of the predicted phase transitions by these approaches beyond the Gaussian matrices has been observed [18]. The RIP analysis is also convenient for bounding the reconstruction error in ℓ_2 norm when the observation noises are present.

Neighborly polytope approach

As discussed earlier, by relating a k -neighborly polytope condition to the success of ℓ_1 minimization for decoding ideally k -sparse signals, Donoho and Tanner gave the precise phase transitions for decoding ideally sparse signals in [15, 20]. The Grassmann angle approach in this chapter is a generalization of the neighborly polytope approach in [15, 20]. Compared with the neighborly polytope condition which only works for analyzing the ideally sparse signal vectors, the generalized Grassmann approach is intended to give sharp phase transitions for the null space “balancedness” conditions, which are useful in a more general setting, for example, in analyzing the robustness of ℓ_1 minimization, weighted ℓ_1 minimization, and iterative reweighted ℓ_1 minimization. Mathematically, in this chapter we need to derive new formulas for the various geometric angles in the Grassmann angle approach. This chapter uses the same computational techniques in estimating the asymptotics of the Grassmann angle as in estimating the asymptotic face counts in [15].

Spherical section property approach

The threshold bounds on ζ for the null space condition to hold was also analyzed in [28, 40, 45, 46], using the *spherical section property* of linear subspaces derived from the Kashin–Garnaev–Gluskin Inequality [23, 27, 46]. The Kashin–Garnaev–Gluskin Inequality claims that for a uniformly distributed $(n - m)$ -dimensional subspace, with overwhelming probability, all vectors w from this subspace will satisfy the spherical section property,

$$\|w\|_1 \geq \frac{c_1 \sqrt{m}}{\sqrt{1 + \log(n/m)}} \|w\|_2, \quad (7.7)$$

where c_1 is a constant independent of the problem dimension. Note that $\|w\|_1 \leq \sqrt{n} \|w\|_2$ always holds, taking equality only if w is a perfectly balanced vector of constant magnitude, so it is natural to see that this spherical section property can be used to investigate the subspace “balancedness” property [45]. This approach extends to general matrices such as random Gaussian matrices, random Bernoulli matrices, and random Fourier mapping matrices [40], and is applicable to the analysis of sparse recovery robustness [46]. The threshold bounds on ζ given by this approach are sometimes better than those obtained from the RIP condition [45], but are generally worse than those obtained by the

neighborly polytope approach, partially because of the coarsely estimated c_1 used in the literature.

Sphere covering approach

The null space condition has also been analyzed by a sphere covering approach in [21, 38, 39]. The subspace property (7.5) is supposed to hold for every vector w in the null space of A , and we can restrict our attention to all the points w in the form of $w = Bv$, where $B \in \mathbb{R}^{n \times (n-m)}$ is a fixed basis for the null space of A , and v is any point from the unit Euclidean sphere in \mathbb{R}^{n-m} . The sphere covering approach proposed to cover the unit Euclidean sphere densely with discrete points such that any point on this unit Euclidean sphere is close enough to a discrete point. If the null space condition (7.5) holds for the vectors generated by these discrete points, it would be possible to infer that the null space condition will also hold for all the points generated by the unit Euclidean sphere and for all the points in the null space of A .

Following this methodology, various threshold bounds have been established in [21, 38, 39]. These bounds are generally better than the threshold bounds from the RIP condition, but weaker than the bounds from the Grassmann angle approach. But in the limiting case when m is very close to n , the threshold bounds from the sphere covering approach can match or are better than the ones obtained from the neighborly polytope approach.

“Escape-through-mesh” approach

More recently, an alternative framework for establishing sharp ℓ_1 minimization thresholds has been proposed in [36] by craftily using the “escape-through-mesh” theorem [24]. The “escape-through-mesh” theorem quantifies the probability that a uniformly distributed $(n-m)$ -dimensional subspace in \mathbb{R}^n misses a set of points on the unit Euclidean sphere in \mathbb{R}^n . The “escape-through-mesh” theorem was first used in analyzing sparse reconstructions in [34]. Based on this theorem, a careful calculation was devised in [36] to evaluate the probability that a uniformly distributed $(n-m)$ -dimensional subspace in \mathbb{R}^n escapes the set of points that violate the null space “balancedness” condition (7.5) for $C = 1$.

The method of [36] yields almost the same threshold bounds for weak recovery that the neighborly polytope approach does; however, for sectional and strong recoveries, it gives different threshold bounds (in some regimes the neighborly polytope approach gives a better bound, and in other regimes the “escape-through-mesh” approach does better). Fully understanding the relation between this “escape-through-mesh” approach and the neighborly polytope approach should be of great interest.

Message passing analysis approach

More recent works [16][17] give threshold bounds for large-scale ℓ_1 minimization and ℓ_1 regularized regression problems through graphical models and novel message passing analysis. For this very interesting approach, the readers are encouraged to refer to Chapter 9 for more details. By comparison, the Grassmann angle approach can provide ℓ_1 norm bounded robustness results in the “weak,” “sectional,” and “strong” senses

(see Section 7.8), while the message passing analysis is more powerful in providing average-case robustness results in terms of mean squared error [17].

7.1.4 Some concepts in high-dimensional geometry

In this part, we will give the explanations of several geometric terminologies often used in this chapter for the purpose of quick reference.

Grassmann manifold

The Grassmann manifold $\text{Gr}_i(j)$ refers to the set of i -dimensional subspaces in the j -dimensional Euclidean space \mathbb{R}^j . It is known that there exists a unique invariant measure μ' on $\text{Gr}_i(j)$ such that $\mu'(\text{Gr}_i(j)) = 1$.

For more facts on the Grassmann manifold, please see [5].

Polytope, face, vertex

A polytope in this chapter refers to the convex hull of a finite number of points in the Euclidean space. Any extreme point of a polytope is a vertex of this polytope. A face of a polytope is defined as the convex hull of a set of its vertices such that no point in this convex hull is an interior point of the polytope. The dimension of a face refers to the dimension of the affine hull of that face. The book [26] offers a nice reference on convex polytopes.

Cross-polytope

The n -dimensional cross-polytope is the polytope of the unit ℓ_1 ball, namely it is the set

$$\{x \in \mathbb{R}^n \mid \|x\|_1 = 1\}.$$

The n -dimensional cross-polytope has $2n$ vertices, namely $\pm e_1, \pm e_2, \dots, \pm e_n$, where e_i , $1 \leq i \leq n$, is the unit vector with its i th coordinate element being 1. Any k extreme points without opposite pairs at the same coordinate will constitute a $(k-1)$ -dimensional face of the cross-polytope. So the cross-polytope will have $2^k \binom{n}{k}$ faces of dimension $(k-1)$.

Grassmann angle

The Grassmann angle for an n -dimensional cone \mathcal{C} under the Grassmann manifold $\text{Gr}_i(n)$, is the measure of the set of i -dimensional subspaces (over $\text{Gr}_i(n)$) which intersect the cone \mathcal{C} nontrivially (namely at some other point besides the origin). For more details on the Grassmann angle, internal angle, and external angle, please refer to [25][26][31].

Cone obtained by observing a set B from a set A

In this chapter, when we say “cone obtained by observing B from A ,” we mean the conic hull of all the vectors in the form of $x_1 - x_2$, where $x_1 \in B$ and $x_2 \in A$.

Internal angle

An internal angle $\beta(F_1, F_2)$, between two faces F_1 and F_2 of a polytope or a polyhedral cone, is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1 . The internal angle $\beta(F_1, F_2)$ is defined to be zero when

$F_1 \not\subseteq F_2$ and is defined to be one if $F_1 = F_2$. Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

External angle

An external angle $\gamma(F_3, F_4)$, between two faces F_3 and F_4 of a polytope or a polyhedral cone, is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3 . The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \not\subseteq F_4$ and is defined to be one if $F_3 = F_4$.

7.1.5 Organization

The rest of the chapter is organized as follows. In Section 7.2, we introduce a null space characterization for guaranteeing robust signal recovery using ℓ_1 minimization. Section 7.3 presents a Grassmann angle-based high-dimensional geometrical framework for analyzing the null space characterization. In Sections 7.4, 7.5, 7.6, and 7.7, analytical performance bounds are given for the null space characterization for matrices that are rotationally invariant, such as those constructed from i.i.d. Gaussian entries. Section 7.8 shows how the Grassmann angle analytical framework can be extended to analyzing the “weak,” “sectional,” and “strong” notations of robust signal recovery.

In Section 7.9, numerical evaluations of the performance bounds for robust signal recovery are given. Section 7.10 and 7.11 will introduce the applications of the Grassmann angle approach to analyzing weighted ℓ_1 minimization and iterative reweighted ℓ_1 minimization algorithms.

Section 7.12 concludes the chapter. In the Appendix, we provide the proofs of related lemmas and theorems.

7.2 The null space characterization

In this section we introduce a useful characterization of the matrix A . The characterization will establish a necessary and sufficient condition on the matrix A so that the solution of (7.2) approximates the solution of (7.1) such that (7.3) holds. (See [22, 30, 45, 14, 38, 39, 28] etc. for variations of this result).

THEOREM 7.2 *Assume that A is a general $m \times n$ measurement matrix. Let $C > 1$ be a positive number. Further, assume that $y = Ax$ and that w is an $n \times 1$ vector. Let K be a subset of $\{1, 2, \dots, n\}$ such that $|K| = k$, where $|K|$ is the cardinality of K and let K_i denote the i th element of K . Further, let $\bar{K} = \{1, 2, \dots, n\} \setminus K$. Then for any $x \in \mathbb{R}^n$, for any K such that $|K| = k$, any solution \hat{x} produced by (7.2) will satisfy*

$$\|x - \hat{x}\|_1 \leq \frac{2(C+1)}{C-1} \|x_{\bar{K}}\|_1, \quad (7.8)$$

if $\forall w \in \mathbb{R}^n$ such that

$$Aw = 0$$

and $\forall K$ such that $|K| = k$, we have

$$C\|w_K\|_1 \leq \|w_{\bar{K}}\|_1. \quad (7.9)$$

Conversely, there exists a measurement matrix A , an x , and corresponding \hat{x} (\hat{x} is a minimizer to the programming (7.2)) such that (7.9) is violated for some set K with cardinality k and some vector w from the null space of A , and

$$\|x - \hat{x}\|_1 > 2 \frac{(C+1)}{C-1} \|x_{\bar{K}}\|_1.$$

Proof. First, suppose the matrix A has the claimed null space property as in (7.9) and we want to prove that any solution \hat{x} satisfies (7.8). Note that the solution \hat{x} of (7.2) satisfies

$$\|\hat{x}\|_1 \leq \|x\|_1,$$

where x is the original signal. Since $A\hat{x} = y$, it easily follows that $w = \hat{x} - x$ is in the null space of A . Therefore we can further write $\|x\|_1 \geq \|x + w\|_1$. Using the triangular inequality for the ℓ_1 norm we obtain

$$\begin{aligned} \|x_K\|_1 + \|x_{\bar{K}}\|_1 &= \|x\|_1 \\ &\geq \|\hat{x}\|_1 = \|x + w\|_1 \\ &\geq \|x_K\|_1 - \|w_K\|_1 + \|w_{\bar{K}}\|_1 - \|x_{\bar{K}}\|_1 \\ &\geq \|x_K\|_1 - \|x_{\bar{K}}\|_1 + \frac{C-1}{C+1} \|w\|_1 \end{aligned}$$

where the last inequality is from the claimed null space property. Relating the head and tail of the inequality chain above,

$$2\|x_{\bar{K}}\|_1 \geq \frac{(C-1)}{C+1} \|w\|_1.$$

Now we prove the second part of the theorem, namely when (7.9) is violated, there exist scenarios where the error performance bound (7.8) fails.

Consider a generic $m \times n$ matrix A' . For each integer $1 \leq k \leq n$, let us define the quantity h_k as the supremum of $\frac{\|w_K\|_1}{\|w_{\bar{K}}\|_1}$ over all such sets K of size $|K| \leq k$ and over all nonzero vectors w in the null space of A' . Let k^* be the biggest k such that $h_k \leq 1$. Then there must be a nonzero vector w' in the null space of A and a set K^* of size k^* , such that

$$\|w'_{K^*}\|_1 = h_{k^*} \|w'_{\bar{K}^*}\|_1.$$

Now we generate a new measurement matrix A by multiplying the portion A'_{K^*} of the matrix A' by h_{k^*} . Then we will have a vector w in the null space of A satisfying

$$\|w_{K^*}\|_1 = \|w_{\overline{K^*}}\|_1.$$

Now we take a signal vector $x = (-w_{K^*}, 0_{\overline{K^*}})$ and claim that $\hat{x} = (0, w_{\overline{K^*}})$ is a minimizer to the programming (7.2). In fact, recognizing the definition of h_{k^*} , we know all the vectors w'' in the null space of the measurement matrix A will satisfy $\|x + w''\|_1 \geq \|x\|_1$. Let us assume that $k^* \geq 2$ and take $K'' \subseteq K^*$ as the index set corresponding to the largest $(k^* - i)$ elements of x_{K^*} in amplitude, where $1 \leq i \leq (k^* - 1)$. From the definition of k^* , it is apparent that $C' = \frac{\|w_{\overline{K''}}\|_1}{\|w_{K''}\|_1} > 1$ since w is nonzero for any index in the set K^* . Let us now take $C = \frac{\|w_{\overline{K''}}\|_1}{\|w_{K''}\|_1} + \epsilon$, where $\epsilon > 0$ is any arbitrarily small positive number. Thus the condition (7.9) is violated for the vector w , the set K'' , and the defined constant C .

Now by inspection, the decoding error is

$$\|x - \hat{x}\|_1 = \frac{2(C' + 1)}{C' - 1} \|x_{K''}\|_1 > \frac{2(C + 1)}{C - 1} \|x_{K''}\|_1,$$

violating the error bound (7.8) (for the set K''). □

Discussion

It should be noted that if the condition (7.9) is true for all the sets K of cardinality k , then

$$2\|x_{\overline{K}}\|_1 \geq \frac{(C - 1)}{C + 1} \|\hat{x} - x\|_1$$

is also true for the set K which corresponds to the k largest (in amplitude) components of the vector x . So

$$2\sigma_k(x)_1 \geq \frac{(C - 1)}{C + 1} \|\hat{x} - x\|_1$$

which exactly corresponds to (7.3). It is an interesting result that, for a particular *fixed* measurement matrix A , the violation of (7.9) for some $C > 1$ does not necessarily imply that the existence of a vector x and a minimizer solution \hat{x} to (7.2) such that the performance guarantee (7.8) is violated. For example, assume $n = 2$ and the null space of the measurement matrix A is a one-dimensional subspace and has the vector $(1, 100)$ as its basis. Then the null space of the matrix A violates (7.9) with $C = 101$ and the set $K = \{1\}$. But a careful examination shows that the biggest possible $\frac{\|x - \hat{x}\|_1}{\|x_{\overline{K}}\|_1}$ ($\|x_{\overline{K}}\|_1 \neq 0$) is equal to $\frac{100+1}{100} = \frac{101}{100}$, achieved by such an x as $(-1, -1)$. In fact, all those vectors $x = (a, b)$ with $b \neq 0$ will achieve $\frac{\|x - \hat{x}\|_1}{\|x_{\overline{K}}\|_1} = \frac{101}{100}$. However, (7.8) has $\frac{2(C+1)}{C-1} = \frac{204}{100}$. This suggests that for a specific measurement matrix A , the tightest error bound for $\frac{\|x - \hat{x}\|_1}{\|x_{\overline{K}}\|_1}$ should involve the detailed structure of the null space of A . But for general measurement matrices A , as suggested by Theorem 7.2, the condition (7.9) is a necessary and sufficient condition to offer the performance guarantee (7.8).

Analyzing the null space condition: the Gaussian ensemble

In the remaining part of this chapter, for a given value $\delta = m/n$ and any value $C \geq 1$, we will devote our efforts to determining the value of feasible $\zeta = \rho\delta = k/n$ for which there exists a sequence of A such that the null space condition (7.9) is satisfied for all the sets K of size k when n goes to infinity and $m/n = \delta$. For a specific A , it is very hard to check whether the condition (7.9) is satisfied or not. Instead, we consider randomly choosing A from the Gaussian ensemble, namely A has i.i.d. $\mathcal{N}(0, 1)$ entries, and analyze for what ζ , the condition (7.9) for its null space is satisfied with overwhelming probability as n goes to infinity. This Gaussian matrix ensemble is widely used in compressive sensing research, see for example, Chapters 1 and 9.

The following lemma gives a characterization of the resulting null space of A , which is a fairly well-known result [8][32].

LEMMA 7.1 *Let $A \in \mathbb{R}^{m \times n}$ be a random matrix with i.i.d. $\mathcal{N}(0, 1)$ entries. Then the following statements hold:*

- *The distribution of A is right-rotationally invariant: for any Θ satisfying $\Theta\Theta^* = \Theta^*\Theta = I$, $P_A(A) = P_A(A\Theta)$.*
- *There exists a basis Z of the null space of A , such that the distribution of Z is left-rotationally invariant: for any Θ satisfying $\Theta\Theta^* = \Theta^*\Theta = I$, $P_Z(Z) = P_Z(\Theta^*Z)$.*
- *It is always possible to choose a basis Z for the null space such that Z has i.i.d. $\mathcal{N}(0, 1)$ entries.*

In view of Theorem 7.2 and Lemma 7.1 what matters is that the null space of A be rotationally invariant. Sampling from this rotationally invariant distribution is equivalent to uniformly sampling a random $(n - m)$ -dimensional subspace from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. For any such A and ideally sparse signals, the sharp bounds of [15] apply. However, we shall see that the neighborly polytope condition for ideally sparse signals does not readily apply to the proposed null space condition analysis for approximately sparse signals, since the null space condition cannot be transformed to the k -neighborly property in a *single* high-dimensional polytope [15]. Instead, in this chapter, we shall give a unified Grassmann angle framework to directly analyze the proposed null space property.

7.3 The Grassmann angle framework for the null space characterization

In this section we detail the Grassmann angle-based framework for analyzing the bounds on $\zeta = k/n$ such that (7.9) holds for every vector in the null space, which we denote by Z . Put more precisely, given a certain constant $C > 1$ (or $C \geq 1$), which corresponds to a certain level of recovery accuracy for the approximately sparse signals, we are interested in what scaling k/n we can achieve while satisfying the following condition on Z ($|K| = k$):

$$\forall w \in Z, \forall K \subseteq \{1, 2, \dots, n\}, C\|w_K\|_1 \leq \|w_K\|_1. \quad (7.10)$$

From the definition of the condition (7.10), there is a tradeoff between the largest sparsity level k and the parameter C . As C grows, clearly the largest k satisfying (7.10) will likely decrease, and, at the same time, ℓ_1 minimization will be more robust in terms of the residual norm $\|x_{\overline{K}}\|_1$. The key in our derivation is the following lemma:

LEMMA 7.2 *For a certain subset $K \subseteq \{1, 2, \dots, n\}$ with $|K| = k$, the event that the null space Z satisfies*

$$C\|w_K\|_1 \leq \|w_{\overline{K}}\|_1, \forall w \in Z$$

is equivalent to the event that $\forall x$ supported on the k -set K (or supported on a subset of K):

$$\|x_K + w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \geq \|x_K\|_1, \forall w \in Z. \quad (7.11)$$

Proof. First, let us assume that $C\|w_K\|_1 \leq \|w_{\overline{K}}\|_1, \forall w \in Z$. Using the triangular inequality, we obtain

$$\begin{aligned} & \|x_K + w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \\ & \geq \|x_K\|_1 - \|w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \\ & \geq \|x_K\|_1 \end{aligned}$$

thus proving the forward part of this lemma. Now let us assume instead that $\exists w \in Z$, such that $C\|w_K\|_1 > \|w_{\overline{K}}\|_1$. Then we can construct a vector x supported on the set K (or a subset of K), with $x_K = -w_K$. Then we have

$$\begin{aligned} & \|x_K + w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \\ & = 0 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \\ & < \|x_K\|_1, \end{aligned}$$

proving the converse part of this lemma. □

Now let us consider the probability that condition (7.10) holds for the sparsity $|K| = k$ if we uniformly sample a random $(n - m)$ -dimensional subspace Z from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$. Based on Lemma 7.2, we can equivalently consider the complementary probability P that there exists a subset $K \subseteq \{1, 2, \dots, n\}$ with $|K| = k$, and a vector $x \in \mathbb{R}^n$ supported on the set K (or a subset of K) failing the condition (7.11). With the linearity of the subspace Z in mind, to obtain P , we can restrict our attention to those vectors x from the cross-polytope (the unit ℓ_1 ball) $\{x \in \mathbb{R}^n \mid \|x\|_1 = 1\}$ that are only supported on the set K (or a subset of K).

First, we upper bound the probability P by a union bound over all the possible support sets $K \subseteq \{1, 2, \dots, n\}$ and all the sign patterns of the k -sparse vector x . Since the k -sparse

vector x has $\binom{n}{k}$ possible support sets of cardinality k and 2^k possible sign patterns (non-negative or non-positive), we have

$$P \leq \binom{n}{k} \times 2^k \times P_{K,-}, \quad (7.12)$$

where $P_{K,-}$ is the probability that for a specific *support set* K , there exists a k -sparse vector x of a specific *sign pattern* which fails the condition (7.11). By symmetry, without loss of generality, we assume the signs of the elements of x to be nonpositive.

So now let us focus on deriving the probability $P_{K,-}$. Since x is a nonpositive k -sparse vector supported on the set K (or a subset of K) and can be restricted to the cross-polytope $\{x \in \mathbb{R}^n \mid \|x\|_1 = 1\}$, x is also on a $(k-1)$ -dimensional face, denoted by F , of the skewed cross-polytope (weighted ℓ_1 ball) SP:

$$\text{SP} = \{y \in \mathbb{R}^n \mid \|y_K\|_1 + \|\frac{y_{\bar{K}}}{C}\|_1 \leq 1\}. \quad (7.13)$$

Then $P_{K,-}$ is the probability that there exists an $x \in F$, and there exists a $w \in Z$ ($w \neq 0$) such that

$$\|x_K + w_K\|_1 + \|\frac{w_{\bar{K}}}{C}\|_1 \leq \|x_K\|_1 = 1. \quad (7.14)$$

We first focus on studying a specific *single* point $x \in F$, without loss of generality, assumed to be in the relative interior of this $(k-1)$ -dimensional face F . For this single particular x on F , the probability, denoted by P'_x , that $\exists w \in Z$ ($w \neq 0$) such that (7.14) holds is essentially the probability that a uniformly chosen $(n-m)$ -dimensional subspace Z shifted by the point x , namely $(Z+x)$, intersects the skewed cross-polytope

$$\text{SP} = \{y \in \mathbb{R}^n \mid \|y_K\|_1 + \|\frac{y_{\bar{K}}}{C}\|_1 \leq 1\} \quad (7.15)$$

nontrivially, namely, at some other point besides x .

From the linear property of the subspace Z , the event that $(Z+x)$ intersects the skewed cross-polytope SP is equivalent to the event that Z intersects nontrivially with the cone $\text{SP-Cone}(x)$ obtained by observing the skewed polytope SP from the point x . (Namely, $\text{SP-Cone}(x)$ is the conic hull of the point set $(\text{SP} - x)$ and $\text{SP-Cone}(x)$ has the origin of the coordinate system as its apex.) However, as noticed in the geometry of convex polytopes [25][26], the $\text{SP-Cone}(x)$ is identical for any x lying in the relative interior of the face F . This means that the probability $P_{K,-}$ is equal to P'_x , regardless of the fact that x is only a single point in the relative interior of the face F . (The acute reader may have noticed some singularities here because $x \in F$ may not be in the relative interior of F , but it turns out that the $\text{SP-Cone}(x)$ is then only a subset of the cone we get when x is in the relative interior of F . So we do not lose anything if we restrict x to be in the relative interior of the face F .) In summary, we have

$$P_{K,-} = P'_x.$$

Now we only need to determine P'_x . From its definition, P'_x is exactly the *complementary Grassmann angle* [25] for the face F with respect to the polytope SP under the Grassmann manifold $\text{Gr}_{(n-m)}(n)$: the probability of a uniformly distributed $(n-m)$ -dimensional subspace Z from the Grassmann manifold $\text{Gr}_{(n-m)}(n)$ intersecting nontrivially with the cone $\text{SP-Cone}(x)$ formed by observing the skewed cross-polytope SP from the relative interior point $x \in F$.

Building on the works by L. A. Santaló [35] and P. McMullen [31] etc. in high-dimensional integral geometry and convex polytopes, the complementary Grassmann angle for the $(k-1)$ -dimensional face F can be explicitly expressed as the sum of products of internal angles and external angles [26]:

$$2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}), \quad (7.16)$$

where s is any non-negative integer, G is any $(m+1+2s)$ -dimensional face of the skewed cross-polytope ($\mathfrak{S}_{m+1+2s}(\text{SP})$ is the set of all such faces), $\beta(\cdot, \cdot)$ stands for the internal angle, and $\gamma(\cdot, \cdot)$ stands for the external angle.

The internal angles and external angles are basically defined as follows [26][31]:

- An internal angle $\beta(F_1, F_2)$ is the fraction of the hypersphere S covered by the cone obtained by observing the face F_2 from the face F_1 .² The internal angle $\beta(F_1, F_2)$ is defined to be zero when $F_1 \not\subseteq F_2$ and is defined to be one if $F_1 = F_2$.
- An external angle $\gamma(F_3, F_4)$ is the fraction of the hypersphere S covered by the cone of outward normals to the hyperplanes supporting the face F_4 at the face F_3 . The external angle $\gamma(F_3, F_4)$ is defined to be zero when $F_3 \not\subseteq F_4$ and is defined to be one if $F_3 = F_4$.

Let us take for example the 2-dimensional skewed cross-polytope

$$\text{SP} = \{(y_1, y_2) \in \mathbb{R}^2 \mid \|y_2\|_1 + \left\| \frac{y_1}{C} \right\|_1 \leq 1\}$$

(namely the diamond) in Figure 7.2, where $n = 2$, $(n-m) = 1$ and $k = 1$. Then the point $x = (0, -1)$ is a 0-dimensional face (namely a vertex) of the skewed polytope SP. Now from their definitions, the internal angle $\beta(x, \text{SP}) = \beta/2\pi$ and the external angle $\gamma(x, \text{SP}) = \gamma/2\pi$, $\gamma(\text{SP}, \text{SP}) = 1$. The complementary Grassmann angle for the vertex x with respect to the polytope SP is the probability that a uniformly sampled 1-dimensional subspace (namely a line, we denote it by Z) shifted by x intersects nontrivially with $\text{SP} = \{(y_1, y_2) \in \mathbb{R}^2 \mid \|y_2\|_1 + \left\| \frac{y_1}{C} \right\|_1 \leq 1\}$ (or equivalently the probability that Z intersects nontrivially with the cone obtained by observing SP from the point x). It is obvious that this probability is β/π . The readers can also verify the correctness of the formula (7.16) very easily for this toy example.

² Note the dimension of the hypersphere S here matches the dimension of the corresponding cone discussed. Also, the center of the hypersphere is the apex of the corresponding cone. All these defaults also apply to the definition of the external angles.

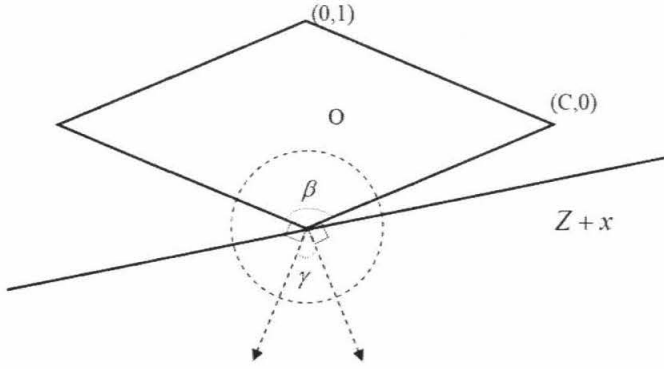


Figure 7.2 The Grassmann angle for a skewed cross-polytope. [42] © 2008 IEEE.

Generally, it might be hard to give explicit formulae for the external and internal angles involved, but fortunately in the skewed cross-polytope case, both the internal angles and the external angles can be explicitly derived. The derivations of these quantities involve the computations of the volumes of cones in high-dimensional geometry and will be presented in the appendix. Here we only present the final results.

Firstly, let us look at the internal angle $\beta(F, G)$ between the $(k-1)$ -dimensional face F and a $(l-1)$ -dimensional face G . Notice that the only interesting case is when $F \subseteq G$ since $\beta(F, G) \neq 0$ only if $F \subseteq G$. We will see if $F \subseteq G$, the cone formed by observing G from F is the direct sum of a $(k-1)$ -dimensional linear subspace and a convex polyhedral cone formed by $(l-k)$ unit vectors with inner product $\frac{1}{1+C^2k}$ between each other. So the internal angle derived in the appendix is given by

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \quad (7.17)$$

where $V_i(S^i)$ denotes the i th-dimensional surface measure on the unit sphere S^i , while $V_i(\alpha', i)$ denotes the surface measure for a regular spherical simplex with $(i+1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i+1)$ vertices. Thus in the appendix, (7.17) is shown to be equal to $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m'-1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta), \quad (7.18)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda. \quad (7.19)$$

Secondly, we derive in the appendix the external angle $\gamma(G, \text{SP})$ between the $(l-1)$ -dimensional face G and the skewed cross-polytope SP as:

$$\gamma(G, \text{SP}) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^{\infty} e^{-x^2} \left(\int_0^{\frac{x}{\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx. \quad (7.20)$$

In summary, combining (7.12), (7.16), (7.17), and (7.20), we get an upper bound on the probability P . If we can show that for a certain $\zeta = k/n$, P goes to zero exponentially in n as $n \rightarrow \infty$, then we know that for such ζ , the null space condition (7.10) holds with overwhelming probability. This is the guideline for computing the bound on ζ in the following sections.

7.4 Evaluating the threshold bound ζ

In summary, we have

$$P \leq \binom{n}{k} \times 2^k \times 2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}). \quad (7.21)$$

This upper bound on the failure probability is similar to the upper bound on the expected number of faces lost in the random projection of the standard ℓ_1 ball through the random projection A , which was originally derived in [1] and used in [15]. However, there are two differences between these two upper bounds. Firstly, different from [15], in (7.21), there do not exist terms dealing with faces F whose dimension is smaller than $(k-1)$. This is because we do not lose anything by only considering the Grassmann angle for a point in the relative interior of a $(k-1)$ -dimensional face F , as explained in the previous section. Secondly, the internal angles and external angles expressions in (7.21) will change as a function of $C \geq 1$, while the corresponding angles in (7.21) are for the neighborly polytopes, where $C = 1$.

In the next few sections, we will build on the techniques developed in [15, 41] to evaluate the bounds on ζ from (7.21) such that P asymptotically goes to 0 as n grows, taking into account the variable $C > 1$. To illustrate the effect of C on the bound ζ , also for the sake of completeness, we will keep the detailed derivations. In the meanwhile, to make the steps easier for the readers to follow, we adopt the same set of notations as in [15] for corresponding quantities.

For simplicity of analysis, we define $l = (m+1+2s)+1$ and $\nu = l/n$. In the skewed cross-polytope SP , we notice that there are in total $\binom{n-k}{l-k} 2^{l-k}$ faces G of dimension $(l-1)$ such that $F \subseteq G$ and $\beta(F, G) \neq 0$. Because of symmetry, it follows from (7.21) that

$$P \leq \underbrace{\sum_{s \geq 0} 2 \binom{n}{k} 2^l \times \binom{n-k}{l-k}}_{D_s} \beta(F, G) \gamma(G, \text{SP}), \quad (7.22)$$

where $l = (m+1+2s)+1$ and $G \subseteq \text{SP}$ is any single face of dimension $(l-1)$ such that $F \subseteq G$. We also define each sum term and its coefficient as D_s and COM_s , as illustrated in (7.22).

In order for the upper bound on P in (7.22) to decrease to 0 as $n \rightarrow \infty$, one sufficient condition is that every sum term D_s in (7.22) goes to 0 exponentially fast in n . Since

$$n^{-1} \log(D_s) = n^{-1} \log(COM_s) + n^{-1} \log(\gamma(G, SP)) + n^{-1} \log(\beta(F, G)),$$

if we want the natural logarithm $n^{-1} \log(D_s)$ to be negative, $n^{-1} \log(COM_s)$, which is non-negative, needs to be overwhelmed by the sum of the logarithms, which are non-positive, for internal angles and external angles.

For fixed ρ , δ , and C , it turns out that there exists a decay exponent $\psi_{ext}(\nu; \rho, \delta, C)$, as a function of $\nu = l/n$, at which rate $\gamma(G, SP)$ decays exponentially. Namely for each $\epsilon > 0$, we have

$$n^{-1} \log(\gamma(G, SP)) \leq -\psi_{ext}(\nu; \rho, \delta, C) + \epsilon,$$

uniformly in $l \geq \delta n$, $n \geq n_0(\rho, \delta, \epsilon, C)$, where $n_0(\rho, \delta, \epsilon, C)$ is a large enough natural number depending only on ρ , δ , ϵ , and C . This exponent $\psi_{ext}(\nu; \rho, \delta, C)$ is explicitly specified in Section 7.6.

Similarly for fixed ρ , δ , and C , the internal angle $\beta(F, G)$ decays at a rate $\psi_{int}(\nu; \rho, \delta, C)$, which is defined in Section 7.5. Namely, for any $\epsilon > 0$, we will have the scaling

$$n^{-1} \log(\beta(F, G)) = -\psi_{int}(\nu; \rho, \delta, C) + \epsilon,$$

uniformly over $l \geq \delta n$ and $n \geq n_0(\rho, \delta, \epsilon, C)$, where $n_0(\rho, \delta, \epsilon, C)$ is a large enough natural number.

For the coefficient term COM_s in (7.21), after some algebra, we know that for any $\epsilon > 0$,

$$n^{-1} \log(COM_s) = \underbrace{\nu \log(2) + H(\rho\delta) + H\left(\frac{\nu - \rho\delta}{1 - \rho\delta}\right)(1 - \rho\delta)}_{\text{combinatorial growth exponent } \psi_{com}(\nu; \rho, \delta)} + \epsilon, \quad (7.23)$$

uniformly when $l \geq \delta n$ and $n > n_0(\rho, \delta, \epsilon)$ (where $n_0(\rho, \delta, \epsilon)$ is some big enough natural number), and $H(p) = p \log(1/p) + (1-p) \log(1/(1-p))$. In getting (7.23), we used the well-known fact that $\frac{1}{n} \log\left(\binom{n}{pn}\right)$ approaches $H(p)$ arbitrarily close as n grows to infinity [15].

In summary, if we define the net exponent $\psi_{net}(\nu; \rho, \delta, C) = \psi_{com}(\nu; \rho, \delta) - \psi_{int}(\nu; \rho, \delta, C) - \psi_{ext}(\nu; \rho, \delta, C)$, then for an arbitrary $C \geq 1$, for any fixed choice of ρ , δ , $\epsilon > 0$, and for large enough n ,

$$n^{-1} \log(D_s) \leq \psi_{net}(\nu; \rho, \delta, C) + 3\epsilon, \quad (7.24)$$

holds uniformly over the sum parameter s in (7.16).

Now we are ready to define the threshold bound $\rho_N(\delta, C)$ such that whenever $\rho < \rho_N(\delta, C)$, the probability P in (7.21) will be decaying to 0 exponentially fast as n grows.

DEFINITION 7.1 For any $\delta \in (0, 1]$, and any $C \geq 1$, we define the critical threshold $\rho_N(\delta, C)$ as the supremum of $\rho \in [0, 1]$ such that for any $\nu \in [\delta, 1]$,

$$\psi_{net}(\nu; \rho, \delta, C) < 0.$$

Now it is time to describe how to calculate the exponents ψ_{int} and ψ_{ext} for the internal angles and external angles respectively. When the parameters ρ , δ , and C are clear from the context, we will omit them from the notations for the combinatorial, internal, and external exponents.

7.5 Computing the internal angle exponent

In this section, we first state how the internal angle exponent $\psi_{int}(\nu; \rho, \delta, C)$ is computed and then justify this computation.

For each ν , we take the function

$$\xi_{\gamma'}(y) = \frac{1 - \gamma'}{\gamma'} y^2 / 2 + \Lambda^*(y), \quad (7.25)$$

where

$$\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}},$$

and $\Lambda^*(\cdot)$ is the dual large deviation rate function given by

$$\Lambda^*(y) = \max_s sy - \Lambda(s).$$

Here $\Lambda(s)$ is the cumulant generating function

$$\Lambda(s) = \log(E(\exp(sY))) = \frac{s^2}{2} + \log(2\Phi(s)),$$

for a standard half-normal random variable Y , where Φ is the cumulative distribution function of a standard Gaussian random variable $\mathcal{N}(0, 1)$. Note that a standard half-normal random variable $Y \sim HN(0, 1)$ is the absolute value $|X|$ of a standard Gaussian random variable $X \sim \mathcal{N}(0, 1)$.

Since the dual large deviation rate function $\Lambda^*(\cdot)$ is a convex function that takes its minimum at $E(Y) = \sqrt{2/\pi}$, so $\xi_{\gamma'}(y)$ is also a convex function that takes its minimum in the interval $(0, \sqrt{2/\pi})$. Let us denote the minimizer of $\xi_{\gamma'}(y)$ as $y_{\gamma'}$. Then for any $C \geq 1$, the internal angle exponent can be computed as

$$\psi_{int}(\nu; \rho, \delta, C) = (\xi_{\gamma'}(y_{\gamma'}) + \log(2))(\nu - \rho\delta). \quad (7.26)$$

Next, we will show (7.26) is indeed the internal angle decay exponent for an arbitrary $C \geq 1$; namely we will prove the following lemma in the same spirit as Lemma 6.1 from [15]:

LEMMA 7.3 *For $k = \rho\delta n$, any $\epsilon > 0$, and any $C \geq 1$, when $n > n_0(\rho, \delta, \epsilon, C)$, where $n_0(\rho, \delta, \epsilon, C)$ is a large enough number,*

$$n^{-1} \log(\beta(F, G)) \leq -\psi_{int}(l/n; \rho, \delta, C) + \epsilon,$$

uniformly for any $l \geq \delta n$.

In fact, using the formula for the internal angle derived in the appendix, we know that

$$-n^{-1} \log(\beta(F, G)) = -n^{-1} \log(B(\frac{1}{1+C^2k}, l-k)), \quad (7.27)$$

where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m'-1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta), \quad (7.28)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda. \quad (7.29)$$

To evaluate (7.27), we need to evaluate the complex integral in $J(m', \theta')$. A saddle point method based on contour integration was sketched for similar integral expressions in [41]. A probabilistic method using large deviation theory for evaluating similar integrals was developed in [15]. Both of these two methods can be applied in our case and of course they will produce the same final results. We will follow the probabilistic method in this chapter. The basic idea is to see the integral in $J(m', \theta')$ as the convolution of $(m' + 1)$ probability densities being expressed in the Fourier domain.

LEMMA 7.4 [15] *Let $\theta = (1 - \alpha')/\alpha'$. Let T be a random variable with the $N(0, \frac{1}{2})$ distribution, and let $W_{m'}$ be a sum of m' i.i.d. half normals $U_i \sim HN(0, \frac{1}{2\theta})$. Let T and $W_{m'}$ be stochastically independent, and let $g_{T+W_{m'}}$ denote the probability density function of the random variable $T + W_{m'}$. Then³*

$$B(\alpha', m') = \sqrt{\frac{\alpha'(m'-1)+1}{1-\alpha'}} \cdot 2^{-m'} \cdot \sqrt{\pi} \cdot g_{T+W_{m'}}(0). \quad (7.30)$$

Here we apply this lemma to $\alpha' = \frac{1}{C^2k+1}$ for general $C \geq 1$. Applying this probabilistic interpretation and large deviation techniques, it is evaluated as in [15] that

$$g_{T+W_{m'}}(0) \leq \frac{2}{\sqrt{\pi}} \cdot \left(\int_0^{\mu_{m'}} v e^{-v^2 - m' \Lambda^*(\frac{\sqrt{2\theta}}{m'} v)} dv + e^{-\mu_{m'}^2} \right), \quad (7.31)$$

where Λ^* is the rate function for the standard half-normal random variable $HN(0, 1)$ and $\mu_{m'}$ is the expectation of $W_{m'}$. In fact, the second term in the sum is negligible because it decays at a greater speed than the first term as the dimension m' grows (to see this, note that $-v^2 - m' \Lambda^*(\frac{\sqrt{2\theta}}{m'} v)$ is a concave function achieving its maximum when $v < \mu_{m'}$; and $-v^2 - m' \Lambda^*(\frac{\sqrt{2\theta}}{m'} v)$ is equal to $-\mu_{m'}^2$, when $v = \mu_{m'}$. Laplace's methods discussed below then show the integral in the first term indeed decays at a slower speed than $e^{-\mu_{m'}^2}$). And after taking $y = \frac{\sqrt{2\theta}}{m'} v$, we have an upper bound for the first term:

$$\frac{2}{\sqrt{\pi}} \cdot \frac{m'^2}{2\theta} \cdot \int_0^{\sqrt{2/\pi}} y e^{-m'(\frac{m'}{2\theta})y^2 - m' \Lambda^*(y)} dy. \quad (7.32)$$

³ In [15], the term $2^{-m'}$ was $2^{1-m'}$, but we believe that $2^{-m'}$ is the right term.

As we know, m' in the exponent of (7.32) is defined as $(l - k)$. Now we notice that the function $\xi_{\gamma'}$ in (7.25) appears in the exponent of (7.32), with $\gamma' = \frac{\theta}{m' + \theta}$. Since $\theta = \frac{1 - \alpha'}{\alpha'} = C^2 k$, we have

$$\gamma' = \frac{\theta}{m' + \theta} = \frac{C^2 k}{(C^2 - 1)k + l}.$$

Since k scales as $\rho \delta n$ and l scales as νn , we further have

$$\gamma' = \frac{k}{\frac{l}{C^2} + \frac{C^2 - 1}{C^2} k} = \frac{\rho \delta}{\frac{C^2 - 1}{C^2} \rho \delta + \frac{\nu}{C^2}},$$

which is apparently consistent with our previous definition of γ' in (7.25).

Recall $\xi_{\gamma'}(y)$ defined in (7.25), then standard Laplace's method will give the upper bound

$$g_{T+W_{m'}}(0) \leq e^{-m' \xi_{\gamma'}(y_{\gamma'})} R_{m'}(\gamma'),$$

where $m'^{-1} \sup_{\gamma' \in [\eta, 1]} \log(R_{m'}(\gamma')) = o(1)$ as $m' \rightarrow \infty$ for any $\eta > 0$.

Plugging this upper bound back into (7.30) and recalling $m' = (\nu - \rho \delta)n$, for any $\epsilon > 0$, with large enough n ,

$$n^{-1} \log(\beta(F, G)) \leq (-\xi_{\gamma'}(y_{\gamma'}) - \log(2))(\nu - \rho \delta) + \epsilon,$$

holds uniformly over $l \geq \nu n$, generalizing the $C = 1$ case in [15].

For any $C \geq 1$, as shown in [15], $\xi_{\gamma'}(y_{\gamma'})$ scales like

$$\frac{1}{2} \log\left(\frac{1 - \gamma'}{\gamma'}\right), \text{ as } \gamma' \rightarrow 0. \quad (7.33)$$

Because $\gamma' = \frac{\rho \delta}{\frac{C^2 - 1}{C^2} \rho \delta + \frac{\nu}{C^2}}$, for any $\nu \in [\delta, 1]$, if we take ρ small enough, γ' can become arbitrarily small. The asymptotic (7.33) means that as $\rho \rightarrow 0$,

$$\psi_{int}(\nu; \rho, \delta, C) \geq \left(\frac{1}{2} \cdot \log\left(\frac{1 - \gamma'}{\gamma'}\right)(1 - \eta) + \log(2)\right)(\nu - \rho \delta). \quad (7.34)$$

This generalizes the $C = 1$ case in [15]. Notice as C increases, the internal angle exponent asymptotic (7.34) decreases.

7.6 Computing the external angle exponent

Closely following [15], let X be a half-normal $HN(0, 1/2)$ random variable, namely a random variable $X = |Z|$ where $Z \sim \mathcal{N}(0, 1/2)$. For $\nu \in (0, 1]$, define x_ν as the solution of

$$\frac{2xG(x)}{g(x)} = \frac{1 - \nu}{\nu'}, \quad (7.35)$$

where $\nu' = (C^2 - 1)\rho\delta + \nu$, $G(x)$ is the cumulative distribution function of X and thus $G(x)$ is the error function

$$G(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy, \quad (7.36)$$

and $g(x) = \frac{2}{\sqrt{\pi}} \exp(-x^2)$ for $x \geq 0$ is the density function for X .

Keeping in mind the dependence of x_ν on $C \geq 1$, we define

$$\psi_{ext}(\nu; \rho, \delta, C) = -(1 - \nu) \log(G(x_\nu)) + \nu' x_\nu^2.$$

When $C = 1$, we have the asymptotic from [15]

$$\psi_{ext}(\nu; \rho, \delta, 1) \sim \nu \log\left(\frac{1}{\nu}\right) - \frac{1}{2} \nu \log(\log(\frac{1}{\nu})) + o(\nu), \nu \rightarrow 0. \quad (7.37)$$

We now set out to prove that the defined external angle exponent is indeed the right exponent. We first give the explicit formula for the external angle formula as a function of the parameter $C \geq 1$ in the appendix. Extracting the exponent from the external angle formula follows [15] and includes the necessary changes to take into account the parameter $C \geq 1$. The justification is summarized in this following lemma:

LEMMA 7.5 For any $C \geq 1$, $\rho = k/n$, and $\delta = m/n$, then for any fixed $\epsilon_1 > 0$,

$$n^{-1} \log(\gamma(G, SP)) < -\psi_{ext}\left(\frac{l}{n}; \rho, \delta, C\right) + \epsilon_1, \quad (7.38)$$

uniformly in $l \geq \delta n$, when n is large enough.

Proof. In the appendix, we derived the explicit integral formula for the external angle:

$$\gamma(G, SP) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k+\frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx. \quad (7.39)$$

After changing integral variables, we have

$$\begin{aligned} \gamma(G, SP) &= \sqrt{\frac{(C^2 - 1)k + l}{\pi}} \\ &\int_0^\infty e^{-((C^2 - 1)k + l)x^2} \left(\frac{2}{\sqrt{\pi}} \int_0^x e^{-y^2} dy \right)^{n-l} dx. \end{aligned} \quad (7.40)$$

Let $\nu = l/n$, $\nu' = (C^2 - 1)\rho\delta + \nu$ then the integral formula can be written as

$$\sqrt{\frac{n\nu'}{\pi}} \int_0^\infty e^{-n\nu'x^2 + n(1-\nu)\log(G(x))} dx, \quad (7.41)$$

where G is the error function from (7.36). To look at the asymptotic behavior of (7.41), following the same methodology as in [15], we first define

$$f_{\rho,\delta,\nu,n}(y) = e^{-n\psi_{\rho,\delta,\nu}(y)} \cdot \sqrt{\frac{n\nu'}{\pi}} \quad (7.42)$$

with

$$\psi_{\rho,\delta,\nu}(y) = \nu' y^2 - (1 - \nu) \log(G(y)).$$

Applying Laplace's method to $\psi_{\rho,\delta,\nu}$ gives Lemma 7.6, which is in the spirit of Lemma 5.2 in [15], and we omit its proof in this chapter.

LEMMA 7.6 For $C \geq 1$ and $\nu \in (0, 1)$, let x_ν denote the minimizer of $\psi_{\rho,\delta,\nu}$. Then

$$\int_0^\infty f_{\rho,\delta,\nu,n}(x) dx \leq e^{-n\psi_{\rho,\delta,\nu}(x_\nu)(1+R_n(\nu))},$$

where for $\delta, \eta > 0$,

$$\sup_{\nu \in [\delta, 1-\eta]} R_n(\nu) = o(1) \text{ as } n \rightarrow \infty,$$

and x_ν is exactly the same x_ν defined earlier in (7.35).

Recall that the defined exponent ψ_{ext} is given by

$$\psi_{ext}(\nu; \rho, \delta, C) = \psi_{\rho,\delta,\nu}(x_\nu). \quad (7.43)$$

From the definition of $\psi_{\rho,\delta,\nu}(x_\nu)$ and (7.43), it is not hard to see that as $\nu \rightarrow 1$, $x_\nu \rightarrow 0$ and $\psi_{ext}(\nu; \rho, \delta, C) \rightarrow 0$. So from (7.43) and Lemma 7.6,

$$n^{-1} \log(\gamma(G, \text{SP})) < -\psi_{ext}(l/n; \rho, \delta, C) + \epsilon_1,$$

uniformly in $l \geq \delta n$, when n is large enough. □

7.7 Existence and scaling of $\rho_N(\delta, C)$

Recall that in determining $\rho_N(\delta, C)$, ψ_{com} is the exponent which must be overwhelmed by the other two exponents $\psi_{int} + \psi_{net}$. The asymptotic relations (7.37) and (7.33) allow us to see the following key facts about $\rho_N(\delta, C)$, the proofs of which are given in the appendix.

LEMMA 7.7 For any $\delta > 0$ and any $C > 1$, we have

$$\rho_N(\delta, C) > 0, \delta \in (0, 1). \quad (7.44)$$

This generalizes the nontriviality of $\rho_N(\delta, C)$ to arbitrary $C \geq 1$. Finally, we have the lower and upper bounds for $\rho_N(\delta, C)$, which shows the scaling bounds for $\rho_N(\delta, C)$ as a function of C .

LEMMA 7.8 When $C \geq 1$, for any fixed $\delta > 0$,

$$\Omega\left(\frac{1}{C^2}\right) \leq \rho_N(\delta, C) \leq \frac{1}{C+1}, \quad (7.45)$$

where $\Omega(\frac{1}{C^2}) \leq \rho_N(\delta, C)$ means that there exists a constant $\iota(\delta)$,

$$\frac{\iota(\delta)}{C^2} \leq \rho_N(\delta, C), \quad \text{as } C \rightarrow \infty,$$

where we can take $\iota(\delta) = \rho_N(\delta, 1)$.

7.8 “Weak,” “sectional,” and “strong” robustness

So far, we have discussed the robustness of ℓ_1 minimization for sparse signal recovery in the “strong” case, namely we required robust signal recovery for all the approximately k -sparse signal vectors x . But in applications or performance analysis, we are also often interested in the signal recovery robustness in weaker senses. As we shall see, the framework given in the previous sections can be naturally extended to the analysis of other notions of robustness for sparse signal recovery, resulting in a coherent analysis scheme. For example, we hope to get a tighter performance bound for a particular signal vector instead of a more general, but looser, performance bound for all the possible signal vectors. In this section, we will present our null space conditions on the matrix A to guarantee the performance of the programming (7.2) in the “weak,” “sectional,” and “strong” senses. Here the robustness in the “strong” sense is exactly the robustness we discussed in the previous sections.

THEOREM 7.3 Let A be a general $m \times n$ measurement matrix, x be an n -element vector, and $y = Ax$. Denote K as a subset of $\{1, 2, \dots, n\}$ such that its cardinality $|K| = k$ and further denote $\bar{K} = \{1, 2, \dots, n\} \setminus K$. Let w denote an $n \times 1$ vector. Let $C > 1$ be a fixed number.

- (Weak Robustness) Given a specific set K and suppose that the part of x on K , namely x_K is fixed. $\forall x_{\bar{K}}$, any solution \hat{x} produced by (7.2) satisfies

$$\|x_K\|_1 - \|\hat{x}_K\|_1 \leq \frac{2}{C-1} \|x_{\bar{K}}\|_1$$

and

$$\|(x - \hat{x})_{\bar{K}}\|_1 \leq \frac{2C}{C-1} \|x_{\bar{K}}\|_1,$$

if and only if $\forall w \in \mathbb{R}^n$ such that $Aw = 0$, we have

$$\|x_K + w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \geq \|x_K\|_1; \quad (7.46)$$

- (Sectional Robustness) Given a specific set $K \subseteq \{1, 2, \dots, n\}$. Then $\forall x \in \mathbb{R}^n$, any solution \hat{x} produced by (7.2) will satisfy

$$\|x - \hat{x}\|_1 \leq \frac{2(C+1)}{C-1} \|x_{\overline{K}}\|_1,$$

if and only if $\forall x' \in \mathbb{R}^n, \forall w \in \mathbb{R}^n$ such that $Aw = 0$,

$$\|x'_K + w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \geq \|x'_K\|_1; \quad (7.47)$$

- (Strong Robustness) If for all possible $K \subseteq \{1, 2, \dots, n\}$, and for all $x \in \mathbb{R}^n$, any solution \hat{x} produced by (7.2) satisfies

$$\|x - \hat{x}\|_1 \leq \frac{2(C+1)}{C-1} \|x_{\overline{K}}\|_1,$$

if and only if $\forall K \subseteq \{1, 2, \dots, n\}, \forall x' \in \mathbb{R}^n, \forall w \in \mathbb{R}^n$ such that $Aw = 0$,

$$\|x'_K + w_K\|_1 + \left\| \frac{w_{\overline{K}}}{C} \right\|_1 \geq \|x'_K\|_1. \quad (7.48)$$

Proof. We will first show the sufficiency of the null space conditions for the various definitions of robustness. Let us begin with the “weak” robustness part. Let $w = \hat{x} - x$ and we must have $Aw = A(\hat{x} - x) = 0$. From the triangular inequality for ℓ_1 norm and the fact that $\|x\|_1 \geq \|x + w\|_1$, we have

$$\begin{aligned} & \|x_K\|_1 - \|x_K + w_K\|_1 \\ & \geq \|w_{\overline{K}} + x_{\overline{K}}\|_1 - \|x_{\overline{K}}\|_1 \\ & \geq \|w_{\overline{K}}\|_1 - 2\|x_{\overline{K}}\|_1. \end{aligned}$$

But the condition (7.46) guarantees that

$$\|w_{\overline{K}}\|_1 \geq C(\|x_K\|_1 - \|x_K + w_K\|_1),$$

so we have

$$\|w_{\overline{K}}\|_1 \leq \frac{2C}{C-1} \|x_{\overline{K}}\|_1,$$

and

$$\|x_K\|_1 - \|\hat{x}_K\|_1 \leq \frac{2}{C-1} \|x_{\overline{K}}\|_1.$$

For the “sectional” robustness, again, we let $w = \hat{x} - x$. Then there must exist an $x' \in \mathbb{R}^n$ such that

$$\|x'_K + w_K\|_1 = \|x'_K\|_1 - \|w_K\|_1.$$

Following the condition (7.47), we have

$$\|w_K\|_1 \leq \left\| \frac{w_{\bar{K}}}{C} \right\|_1.$$

Since

$$\|x\|_1 \geq \|x + w\|_1,$$

following the proof of Theorem 7.2, we have

$$\|x - \hat{x}\|_1 \leq \frac{2(C+1)}{C-1} \|x_{\bar{K}}\|_1.$$

The sufficiency of the condition (7.48) for strong robustness also follows.

Necessity: Since in the proof of the sufficiency, equalities can be achieved in the triangular equalities, the conditions (7.46), (7.47), and (7.48) are also necessary conditions for the respective robustness to hold for every x (otherwise, for certain x 's, there will be $x' = x + w$ with $\|x'\|_1 < \|x\|_1$ which violates the respective robustness definitions. Also, such x' can be the solution to (7.2)). The detailed arguments will similarly follow the proof of the second part of Theorem 7.2. \square

The conditions for “weak,” “sectional,” and “strong” robustness seem to be very similar, and yet there are key differences. The “weak” robustness condition is for x with a specific x_K on a specific subset K , the “sectional” robustness condition is for x with arbitrary value x_K on a specific subset K , and the “strong” robustness condition is for x 's with arbitrary value x_K on all possible subsets. Basically, the “weak” robustness condition (7.46) guarantees that the ℓ_1 norm of \hat{x}_K is not too far away from the ℓ_1 norm of x_K and the error vector $w_{\bar{K}}$ scales linearly in ℓ_1 norm as a function of $\|x_{\bar{K}}\|_1$. Notice that if we define

$$\kappa = \max_{Aw=0, w \neq 0} \frac{\|w_K\|_1}{\|w_{\bar{K}}\|_1},$$

then

$$\|x - \hat{x}\|_1 \leq \frac{2C(1+\kappa)}{C-1} \|x_{\bar{K}}\|_1.$$

That means, if κ is not ∞ for a measurement matrix A , $\|x - \hat{x}\|_1$ also approaches 0 when $\|x_{\bar{K}}\|_1$ approaches 0. Indeed, it is not hard to see that, for a given matrix A , $\kappa < \infty$ as long as the rank of matrix A_K is equal to $|K| = k$, which is generally satisfied for $k < m$.

While the “weak” robustness condition is only for one specific signal x , the “sectional” robustness condition instead guarantees that given *any* approximately k -sparse signal mainly supported on the subset K , the ℓ_1 minimization gives a solution \hat{x} close to the original signal by satisfying (7.3). When we measure an approximately k -sparse signal x (the support of the k largest-magnitude components is fixed though unknown to the decoder) using a randomly generated measurement matrix A , the “sectional” robustness

conditions characterize the probability that the ℓ_1 minimization solution satisfies (7.3) for *any* signals for the set K . If that probability goes to 1 as $n \rightarrow \infty$ for any subset K , we know that there exist measurement matrices A 's that guarantee (7.3) on “almost all” support sets (namely, (7.3) is “almost always” satisfied). The “strong” robustness condition instead guarantees the recovery for approximately sparse signals mainly supported on *any* subset K . The “strong” robustness condition is useful in guaranteeing the decoding bound *simultaneously* for *all* approximately k -sparse signals under a single measurement matrix A .

REMARK: We should mention that from a practical point of view weak robustness is the most meaningful and is what can be observed in simulations (since it is impossible to check all x_K and all subsets K to check for sectional and strong robustness).

As expected, after we take $C = 1$ and let (7.46), (7.47), and (7.48) take strict inequality for all $w \neq 0$ in the null space of A , the conditions (7.46), (7.47), and (7.48) are also sufficient and necessary conditions for unique exact recovery of ideally k -sparse signals in “weak,” “sectional,” and “strong” senses [15].

For a given value $\delta = m/n$ and any value $C \geq 1$, we will determine the value of feasible $\zeta = k/n$ for which there exist a sequence of A 's such that these three conditions are satisfied when $n \rightarrow \infty$ and $m/n = \delta$. As manifested by the statements of the three conditions (7.46), (7.47), and (7.48) and the previous discussions in Section 7.3, we can naturally extend the Grassmann angle approach to analyze the bounds for the probabilities that (7.46), (7.47), and (7.48) fail. Here we will denote these probabilities as P_1 , P_2 , and P_3 , respectively. Note that there are $\binom{n}{k}$ possible support sets K and there are 2^k possible sign patterns for signal x_K . From previous discussions, we know that the event that the condition (7.46) fails is the same for all x_K 's of a specific support set and a specific sign pattern. Then following the same line of reasoning as in Section 7.3, we have $P_1 = P_{K,-}$, $P_2 \leq 2^k \times P_1$, and $P_3 \leq \binom{n}{k} \times 2^k \times P_1$, where $P_{K,-}$ is the probability as in (7.12).

It is worthwhile mentioning that the formula for P_1 is exact since there is no union bound involved and so the threshold bound for the “weak” robustness is tight. In summary, the results in this section suggest that even if k is very close to the weak threshold for ideally sparse signals, we can still have robustness results for approximately sparse signals while the results using restricted isometry conditions [10] suggest smaller sparsity level for recovery robustness. This is the first such result.

7.9 Numerical computations on the bounds of ζ

In this section, we will numerically evaluate the performance bounds on $\zeta = k/n$ such that the conditions (7.9), (7.46), (7.47), and (7.48) are satisfied with overwhelming probability as $n \rightarrow \infty$.

First, we know that the condition (7.9) fails with probability

$$P \leq \binom{n}{k} \times 2^k \times 2 \times \sum_{s \geq 0} \sum_{G \in \mathfrak{S}_{m+1+2s}(\text{SP})} \beta(F, G) \gamma(G, \text{SP}). \quad (7.49)$$

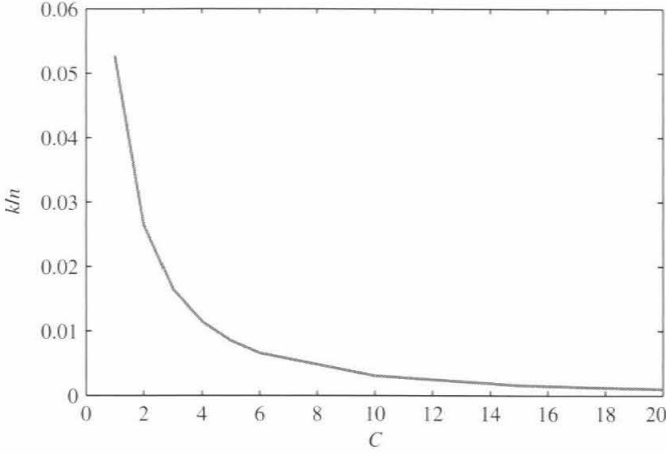


Figure 7.3 Allowable sparsity as a function of C (allowable imperfection of the recovered signal is $\frac{2(C+1)\sigma_k(x)_1}{C-1}$).

Recall that we assume $m/n = \delta$, $l = (m + 1 + 2s) + 1$ and $\nu = l/n$. In order to make P overwhelmingly converge to zero as $n \rightarrow \infty$, following the discussions in Section 7.4, one sufficient condition is to make sure that the exponent for the combinatorial factors

$$\psi_{com} = \lim_{n \rightarrow \infty} \frac{\log \left(\binom{n}{k} 2^k 2^{\binom{n-k}{l-k}} 2^{l-k} \right)}{n} \quad (7.50)$$

and the negative exponent for the angle factors

$$\psi_{angle} = - \lim_{n \rightarrow \infty} \frac{\log (\beta(F, G) \gamma(G, SP))}{n} \quad (7.51)$$

satisfy $\psi_{com} - \psi_{angle} < 0$ uniformly over $\nu \in [\delta, 1)$.

Following [15] we take $m = 0.5555n$. By analyzing the decaying exponents of the external angles and internal angles through the Laplace methods as in Section 7.6, and 7.5, we can compute the numerical results as shown in Figure 7.3, Figure 7.5, and Figure 7.6. In Figure 7.3, we show the largest sparsity level $\zeta = k/n$ (as a function of C) which makes the failure probability of the condition (7.11) approach zero asymptotically as $n \rightarrow \infty$. As we can see, when $C = 1$, we get the same bound $\zeta = 0.095 \times 0.5555 \approx 0.0528$ as obtained for the “weak” threshold for the ideally sparse signals in [15]. As expected, as C grows, the ℓ_1 minimization requires a smaller sparsity level ζ to achieve higher signal recovery accuracy.

In Figure 7.4(a), we show the exponents ψ_{com} , ψ_{int} , ψ_{ext} under the parameters $C = 2$, $\delta = 0.5555$ and $\zeta = 0.0265$. For the same set of parameters, in Figure 7.4(b), we compare the exponents ψ_{com} and ψ_{angle} : the solid curve denotes ψ_{angle} and the dashed curve denotes ψ_{com} . It shows that, under $\zeta = 0.0265$, $\psi_{com} - \psi_{angle} < 0$ uniformly over $\delta \leq \nu \leq 1$. Indeed, $\zeta = 0.0265$ is the bound shown in Figure 7.3 for $C = 2$. In Figure 7.5, for the parameter $\delta = 0.5555$, we give the bounds ζ as a function of C for satisfying

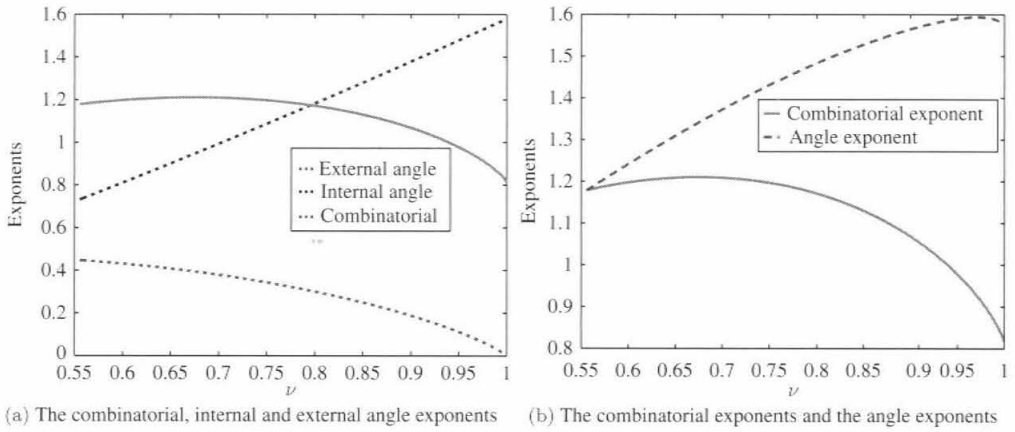


Figure 7.4 The combinatorial, external and internal angle exponent. (a) The combinatorial interref and External angle exponents. (b) The combinatorial exponents and angle exponents.

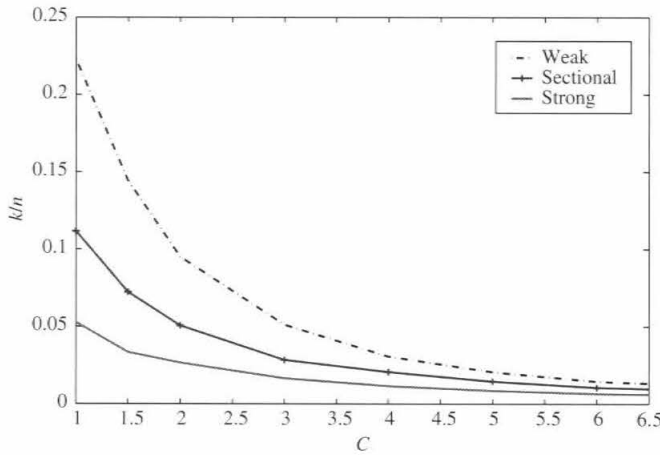


Figure 7.5 The “weak,” “sectional,” and “strong” robustness bounds.

the signal recovery robustness conditions (7.46), (7.47), and (7.48), respectively in the “weak,” “sectional,” and “strong” senses. In Figure 7.6, fixing $C = 2$, we plot how large $\rho = \zeta/\delta$ can be for different δ ’s while satisfying the signal recovery robustness conditions (7.46), (7.47), and (7.48), respectively in “weak,” “sectional,” and “strong” senses.

7.10 Recovery thresholds for weighted ℓ_1 minimization

So far, we have used a null space Grassmann angle geometric approach to give sharp characterizations for the sparsity and ℓ_1 recovery stability tradeoff in compressive sensing. It turns out that the null space Grassmann angle approach is a general framework

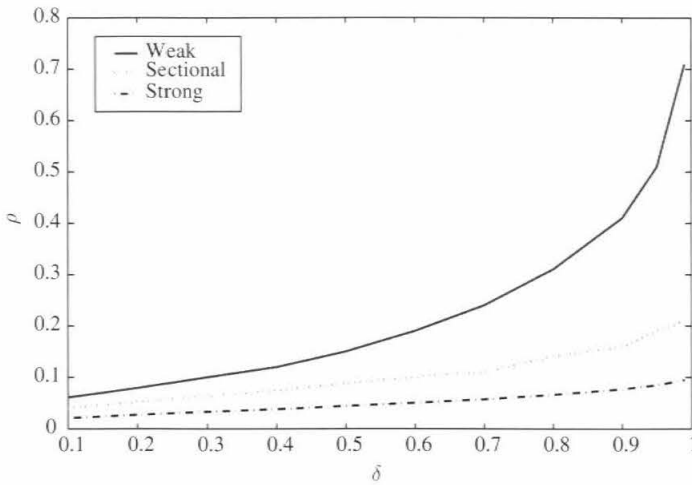


Figure 7.6 The “weak,” “sectional,” and “strong” robustness bounds.

which can be used to give sharp performance bounds for other sparse recovery algorithms, for example, weighted ℓ_1 minimization algorithms and iterative reweighted ℓ_1 algorithms. In these applications, the success of these algorithms can also be reduced to the event that the null space of the measurement matrix intersects trivially with different polyhedral cones. So similarly for these applications, we will be able to characterize the sharp sparsity transition thresholds and in turn, these threshold results will help us optimize the configurations of the weighted algorithms.

The conventional approach to compressive sensing assumes no prior information on the unknown signal vector other than the fact that it is sufficiently sparse over a particular basis. In many applications, however, additional prior information is available, such as in natural images, medical imaging, and in DNA microarrays. How to exploit the structure information in the sparse signals has led to the development of structured sparse models in recent years, see Chapter 1. In the DNA microarrays applications, for instance, signals are often *block sparse*, i.e., the signal is more likely to be nonzero in certain blocks rather than in others [37]. Even when no prior information is available, the preprocessing phases of some sparse recovery algorithms feed “prior” information on the sparse signal (e.g., its sparsity pattern) to the inner-loops of these sparse recovery algorithms [12, 29].

In [29] we consider a particular model for the sparse signal where the entries of the unknown vector fall into a number u of classes, with each class having a specific fraction of nonzero entries. The standard compressed sensing model is therefore a special case where there is only one class. We will focus on the case where the entries of the unknown signal fall into a fixed number u of categories; in the i th set K_i with cardinality n_i , the fraction of nonzero entries is p_i . This model is rich enough to capture many of the salient features regarding prior information. We refer to the signals generated based on this model as *nonuniform sparse* signals. For completeness, we present a general definition.

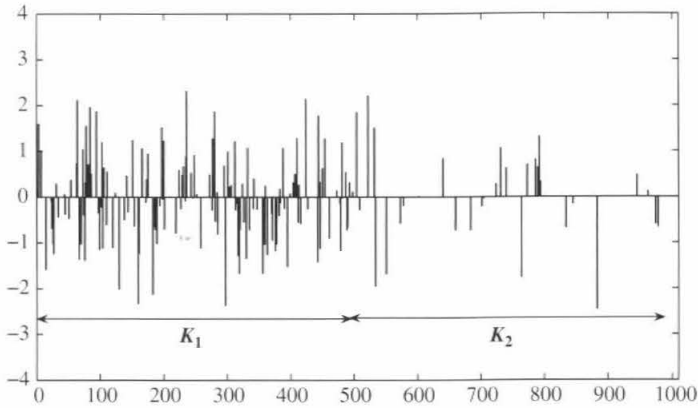


Figure 7.7 Illustration of a nonuniform sparse signal. [29] © 2009 IEEE.

DEFINITION 7.2 Let $\mathcal{K} = \{K_1, K_2, \dots, K_u\}$ be a partition of $\{1, 2, \dots, n\}$, i.e. $(K_i \cap K_j = \emptyset \text{ for } i \neq j, \text{ and } \bigcup_{i=1}^u K_i = \{1, 2, \dots, n\})$, and $P = \{p_1, p_2, \dots, p_u\}$ be a set of positive numbers in $[0, 1]$. An $n \times 1$ vector $x = (x_1, x_2, \dots, x_n)^T$ is said to be a random **nonuniform sparse** vector with sparsity fraction p_i over the set K_i for $1 \leq i \leq u$, if x is generated from the following random procedure:

- Over each set K_i , $1 \leq i \leq u$, the set of nonzero entries of x is a random subset of size $p_i |K_i|$. In other words, a fraction p_i of the entries are nonzero in K_i . p_i is called the sparsity fraction over K_i . The values of the nonzero entries of x can be arbitrary nonzero real numbers.

In Figure 7.7, a sample nonuniform sparse signal with Gaussian distribution for nonzero entries is plotted. The number of sets is considered to be $u = 2$ and both classes have the same size $n/2$, with $n = 1000$. The sparsity fraction for the first class K_1 is $p_1 = 0.3$, and for the second class K_2 is $p_2 = 0.05$.

To accommodate the prior information, one can simply think of modifying ℓ_1 minimization to a weighted ℓ_1 minimization as follows:

$$\min_{Az=y} \|z\|_{w,1} = \min_{Az=y} \sum_{i=1}^n w_i |z_i|. \quad (7.52)$$

The index w on the norm is an indication of the $n \times 1$ non-negative weight vector. Naturally, if we want to suppress the i th entry to be zero in the decoding result, we would like to assign a bigger value to w_i . To boost the performance of sparse recovery, it may benefit to give bigger weights to the blocks where there are more zero elements. For example, in Figure 7.7, we can assign weight $W_1 = 1$ to the first block K_1 and assign another weight $W_2 > 1$ to the sparser block K_2 .

Now the question is, what are the optimal sets of weights for weighted ℓ_1 minimization (7.52) to minimize the number of measurements (or the threshold on the number of

measurements) ensuring a signal vector of the nonuniform sparse model is recovered with overwhelming probability?

This seems to be a very different problem from the ℓ_1 minimization robustness problem we have considered earlier in this chapter. However, these two problems are connected through the null space conditions for the measurement matrix A , and so the Grassmann angle approach in the earlier work can also be applied to this problem. More explicitly, suppose K is the support of a signal vector from the nonuniform sparse model and \bar{K} is the complement of the support set, then the weighted ℓ_1 minimization succeeds in recovering all the vectors supported on K if and only if

$$\|v_K\|_{w_K,1} < \|v_{\bar{K}}\|_{w_{\bar{K}},1} \quad (7.53)$$

holds for every nonzero vector v from the null space of A . The proof of this weighted null space condition is relatively obvious following the same reasoning as in the proof of Theorem 7.2.

In studying this weighted null space condition, one can then extend the Grassmann angle framework to analyze the “failure” probability that the null space of a random A intersects nontrivially with the “weighted” cone of vectors v satisfying

$$\|v_K\|_{w_K,1} \geq \|v_{\bar{K}}\|_{w_{\bar{K}},1}. \quad (7.54)$$

As in the analysis for ℓ_1 minimization robustness, this “failure” probability can be reduced to studying the event that the null space of a random A intersects with a union of “weighted” polyhedral cones. This of course reduces to the computation and evaluation of Grassmann angles for individual cones, only this time for “weighted” polyhedral cones. In fact, for any set of specialized block and sparsity parameters, and for any particular set of weights, one can compute via the Grassmann angle approach the threshold for $\delta_c = m/n$ (the number of measurements needed) such that a sparse signal vector from the nonuniform sparse signal model is recovered with high probability. The derivations and calculations follow the same steps as in previous sections for ℓ_1 minimization robustness, and we will omit them here for the sake of space. For the technical details, the reader is encouraged to read [29]. The main result is stated in the following theorem and its proof can be found in [29].

THEOREM 7.4 *Let $\delta = m/n$ for the Gaussian measurement matrix $A \in \mathbb{R}^{m \times n}$, $\gamma_1 = n_1/n$ and $\gamma_2 = n_2/n$. For fixed values of $\gamma_1, \gamma_2, p_1, p_2, \omega = w_{K_2}/w_{K_1}$, define E to be the event that a random nonuniform sparse vector x_0 (Definition 7.2) with sparsity fractions p_1 and p_2 over the sets K_1 and K_2 respectively with $|K_1| = \gamma_1 n$ and $|K_2| = \gamma_2 n$ is recovered via the weighted ℓ_1 minimization. There exists a computable critical threshold $\delta_c = \delta_c(\gamma_1, \gamma_2, p_1, p_2, \omega)$ such that if $\delta = m/n \geq \delta_c$, then E happens with overwhelming probability as $n \rightarrow \infty$.*

Let us again look at the sample nonuniform sparse signal model in Figure 7.7. For $u = 2$, $\gamma_1 = |K_1|/n = 0.5$, $\gamma_2 = |K_2|/n = 0.5$, $p_1 = 0.3$, and $p_2 = 0.05$, we have numerically computed $\delta_c(\gamma_1, \gamma_2, p_1, p_2, w_{K_2}/w_{K_1})$ as a function of w_{K_2}/w_{K_1} and depicted the

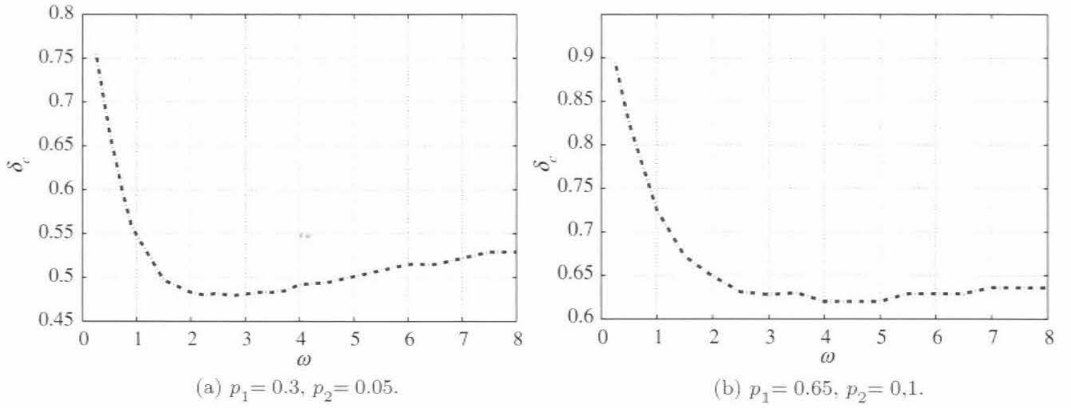


Figure 7.8 δ_c as a function of $\omega = w_{K_2}/w_{K_1}$ for $\gamma_1 = \gamma_2 = 0.5$. [29] © 2009 IEEE.

resulting curve in Figure 7.8(a). This suggests that $w_{K_2}/w_{K_1} \approx 2.5$ is the optimal ratio that one can choose. The value of δ_c for another choice of p_1, p_2 is shown in Figure 7.8(b).

7.11 Approximate support recovery and iterative reweighted ℓ_1

Despite its simplicity and extensive research on other polynomial-complexity sparse recovery algorithms, when no prior information is available, regular ℓ_1 minimization still has the best theoretically established sparse recovery threshold performance for decoding general sparse signal vectors. However, using the Grassmann angle analysis, even when no prior information is available, we are able to show for the first time that a class of (iterative) reweighted ℓ_1 minimization algorithms have strictly higher recovery thresholds on recoverable sparsity levels than regular ℓ_1 minimization, for certain classes of signal vectors whose nonzero elements have fast decaying amplitudes. The technical details of this claim are not presented here due to space limitations, and a more comprehensive study on this can be found in [44].

The reweighted ℓ_1 recovery algorithm proposed in [44] is composed of two steps. In the first step a standard ℓ_1 minimization is used to decode the signal. Note that when the number of nonzero elements is above the recovery threshold, ℓ_1 minimization generally will not give the original sparse signal. Based on ℓ_1 minimization output, a set of entries where the nonzero elements are more likely to reside (the so-called approximate support) are identified. The elements of the unknown signal are thus divided into two classes: one is the approximate support with a relatively higher density of nonzero entries, and the other one is the complement of the approximate support, which has a smaller density of nonzero entries. This corresponds to a nonuniform sparse model in the previous section. The second step of reweighted ℓ_1 recovery is then to perform a weighted ℓ_1 minimization (see the previous section) where elements outside the approximate support set are penalized with a weight larger than 1.

The algorithm is then given as follows, where k is the number of nonzero elements and ω is a weighting parameter which can be adjusted.

Algorithm 7.1

1. Solve the ℓ_1 minimization problem:

$$\hat{x} = \arg \min \|z\|_1 \text{ subject to } Az = Ax. \quad (7.55)$$

2. Obtain an approximation for the support set of x : find the index set $L \subset \{1, 2, \dots, n\}$ which corresponds to the largest k elements of \hat{x} in magnitude.
3. Solve the following weighted ℓ_1 minimization problem and declare the solution as output:

$$x^* = \arg \min \|z_L\|_1 + \omega \|z_{\bar{L}}\|_1 \text{ subject to } Az = Ax. \quad (7.56)$$

For a given number of measurements, if the support size of x , namely $k = |K|$, is slightly larger than the sparsity threshold of ℓ_1 minimization, then the robustness of ℓ_1 minimization, as analyzed via the Grassmann angle approach in this chapter, helps find a lower bound for $\frac{|L \cap K|}{|L|}$, i.e. the density of nonzero elements of x over the set L . With the help of this type of “prior” information about the support of x , the weighted ℓ_1 algorithm, as analyzed via the Grassmann angle approach in the previous section, can be shown to guarantee a full recovery of the original sparse vector even though the number of its nonzero elements is beyond the ℓ_1 minimization recovery threshold, and, at the beginning, we do not have prior support information.

It should be noted that, at the cost of not having any prior information at the beginning of this algorithm, the sparse recovery threshold improvement is not universal over all types of signals. For example, if the nonzero elements of the signal are of a constant amplitude, the support estimate in the first step can be very misleading [12] and leads to bad recovery performance in the second step.

Other variations of reweighted ℓ_1 minimization are given in the literature. For example the algorithm in [12] assigns a different weight to each single entry based on the inverse of the absolute value of its decoding result in regular ℓ_1 minimization, \hat{x} . In some sense, the theoretical results in [44], via the Grassmann angle analysis, explain the threshold improvements observed empirically in [12].

7.12 Conclusion

In this chapter we analyzed a null space characterization of the measurement matrix to guarantee a specific performance for ℓ_1 -norm optimization for approximately sparse

signals. Using high-dimensional geometry tools, we give a unified *null space Grassmann angle*-based analytical framework for compressive sensing. This new framework gives sharp quantitative tradeoffs between the signal sparsity parameter and the recovery accuracy of the ℓ_1 optimization for general signals or approximately sparse signals. As expected, the neighborly polytopes result of [15] for ideally sparse signals can be viewed as a special case on this tradeoff curve. It can therefore be of practical use in applications where the underlying signal is not ideally sparse and where we are interested in the quality of the recovered signal. For example, using the results and their extensions in this chapter and [15], we are able to give a precise sparsity threshold analysis for weighted ℓ_1 minimization when prior information about the signal vector is available [29]. In [44], using the robustness result from this chapter, we are able to show that a two-step weighted ℓ_1 minimization algorithm can provably improve over the sparsity threshold of ℓ_1 minimization for interesting classes of signals, even when prior information is not available.

In essence, this work investigates the fundamental “balancedness” property of linear subspaces, and may be of independent mathematical interest. In future work, it will be interesting to obtain more accurate analysis for compressive sensing under noisy measurements than presented in the current chapter.

7.13 Appendix

7.13.1 Derivation of the internal angles

LEMMA 7.9 Suppose that F is a $(k-1)$ -dimensional face of the skewed cross-polytope

$$SP = \{y \in \mathbb{R}^n \mid \|y_K\|_1 + \|\frac{y_{\bar{K}}}{C}\|_1 \leq 1\}$$

supported on the subset K with $|K| = k$. Then the internal angle $\beta(F, G)$ between the $(k-1)$ -dimensional face F and a $(l-1)$ -dimensional face G ($F \subseteq G$, $G \neq SP$) is given by

$$\beta(F, G) = \frac{V_{l-k-1}(\frac{1}{1+C^2k}, l-k-1)}{V_{l-k-1}(S^{l-k-1})}, \quad (7.57)$$

where $V_i(S^i)$ denotes the i th dimensional surface measure on the unit sphere S^i , while $V_i(\alpha', i)$ denotes the surface measure for a regular spherical simplex with $(i+1)$ vertices on the unit sphere S^i and with inner product as α' between these $(i+1)$ vertices. Equation (7.57) is equal to $B(\frac{1}{1+C^2k}, l-k)$, where

$$B(\alpha', m') = \theta^{\frac{m'-1}{2}} \sqrt{(m'-1)\alpha' + 1} \pi^{-m'/2} \alpha'^{-1/2} J(m', \theta) \quad (7.58)$$

with $\theta = (1 - \alpha')/\alpha'$ and

$$J(m', \theta) = \frac{1}{\sqrt{\pi}} \int_{-\infty}^{\infty} \left(\int_0^{\infty} e^{-\theta v^2 + 2iv\lambda} dv \right)^{m'} e^{-\lambda^2} d\lambda. \quad (7.59)$$

Proof. Without loss of generality, assume that F is a $(k-1)$ -dimensional face with k vertices as $e_p, 1 \leq p \leq k$, where e_p is the n -dimensional standard unit vector with the p th element as "1"; and also assume that the $(l-1)$ -dimensional face G be the convex hull of the l vertices: $e_p, 1 \leq p \leq k$ and $Ce_p, (k+1) \leq p \leq l$. Then the cone $\text{Con}_{F,G}$ formed by observing the $(l-1)$ -dimensional face G of the skewed cross-polytope SP from an interior point x^F of the face F is the positive cone of the vectors:

$$Ce_j - e_i, \text{ for all } j \in J \setminus K, i \in K, \quad (7.60)$$

and also the vectors

$$e_{i_1} - e_{i_2}, \text{ for all } i_1 \in K, i_2 \in K, \quad (7.61)$$

where $J = \{1, 2, \dots, l\}$ is the support set for the face G .

So the cone $\text{Con}_{F,G}$ is the direct sum of the linear hull $L_F = \text{lin}\{F - x^F\}$ formed by the vectors in (7.61) and the cone $\text{Con}_{F^\perp, G} = \text{Con}_{F,G} \cap L_F^\perp$, where L_F^\perp is the orthogonal complement to the linear subspace L_F . Then $\text{Con}_{F^\perp, G}$ has the same spherical volume as $\text{Con}_{F,G}$.

Now let us analyze the structure of $\text{Con}_{F^\perp, G}$. We notice that the vector

$$e_0 = \sum_{p=1}^k e_p$$

is in the linear space L_F^\perp and is also the only such vector (up to linear scaling) supported on K . Thus a vector x in the positive cone $\text{Con}_{F^\perp, G}$ must take the form

$$-\sum_{i=1}^k b_i \times e_i + \sum_{i=k+1}^l b_i \times e_i, \quad (7.62)$$

where $b_i, 1 \leq i \leq l$ are non-negative real numbers and

$$\begin{aligned} C \sum_{i=1}^k b_i &= \sum_{i=k+1}^l b_i, \\ b_1 &= b_2 = \dots = b_k. \end{aligned}$$

That is to say, the $(l-k)$ -dimensional $\text{Con}_{F^\perp, G}$ is the positive cone of $(l-k)$ vectors a^1, a^2, \dots, a^{l-k} , where

$$a^i = C \times e_{k+i} - \sum_{p=1}^k e_p / k, \quad 1 \leq i \leq (l-k).$$

The normalized inner products between any two of these $(l-k)$ vectors is

$$\frac{\langle a^i, a^j \rangle}{\|a^i\| \|a^j\|} = \frac{k \times \frac{1}{k^2}}{C^2 + k \times \frac{1}{k^2}} = \frac{1}{1 + kC^2}.$$

(In fact, a^i 's are also the vectors obtained by observing the vertices e_{k+1}, \dots, e_l from $E_C = \sum_{p=1}^k e_p/k$, the epicenter of the face F .)

We have so far reduced the computation of the internal angle to evaluating (7.57), the relative spherical volume of the cone $\text{Con}_{F^\perp, G}$ with respect to the sphere surface S^{l-k-1} . This was computed as given in this lemma [41, 6] for the positive cones of vectors with equal inner products by using a transformation of variables and the well-known formula

$$V_{i-1}(S^{i-1}) = \frac{i\pi^{\frac{i}{2}}}{\Gamma(\frac{i}{2} + 1)},$$

where $\Gamma(\cdot)$ is the usual Gamma function. □

7.13.2 Derivation of the external angles

LEMMA 7.10 Suppose that F is a $(k-1)$ -dimensional face of the skewed cross-polytope

$$SP = \{y \in \mathbb{R}^n \mid \|y_K\|_1 + \|\frac{y_{\bar{K}}}{C}\|_1 \leq 1\}$$

supported on a subset K with $|K| = k$. Then the external angle $\gamma(G, SP)$ between a $(l-1)$ -dimensional face G ($F \subseteq G$) and the skewed cross-polytope SP is given by

$$\gamma(G, SP) = \frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{C\sqrt{k+\frac{l-k}{C^2}}} e^{-y^2} dy \right)^{n-l} dx. \quad (7.63)$$

Proof. We take the same proof technique of transforming external angle calculation into the integral of Gaussian distribution over the outward normal cone [3]. Without loss of generality, we assume $K = \{1, \dots, k\}$. Since the $(l-1)$ -dimensional face G is the convex hull of k regular vertices with length 1 and $(l-k)$ vertices of length C , without of loss of generality, again we have the $(l-1)$ -dimensional face

$$G = \text{conv}\{e^1, \dots, e^k, C \times e^{k+1}, \dots, C \times e^l\}$$

of the skewed cross-polytope SP . Since there are 2^{n-l} facets containing the face G , the 2^{n-l} outward normal vectors of the supporting hyperplanes of the facets containing G are given by

$$\left\{ \sum_{p=1}^k e_p + \sum_{p=l+1}^n j_p e_p / C + \sum_{p=k+1}^l e_p / C, j_p \in \{-1, 1\} \right\}.$$

The outward normal cone $c(G, SP)$ at the face G is the positive hull of these normal vectors. We also have

$$\begin{aligned} \int_{c(G, SP)} e^{-\|x\|^2} dx &= \gamma(G, SP) V_{n-l}(S^{n-l}) \\ &\times \int_0^\infty e^{-r^2} r^{n-l} dr = \gamma(G, SP) \cdot \pi^{(n-l+1)/2}, \end{aligned} \quad (7.64)$$

where $V_{n-l}(S^{n-l})$ is the spherical volume of the $(n-l)$ -dimensional sphere S^{n-l} .

Suppose a vector in the cone $c(G, \text{SP})$ takes value t at index $i = 1$, then that vector can take any value in the interval $[-t/C, t/C]$ at those indices $(l+1) \leq i \leq n$ (due to the negative and positive signs of the outward normal vectors at those indices) and that vector must take the value t/C at the indices $(k+1) \leq i \leq l$. So we only need $(n-l+1)$ free variables in describing the outward normal cone $c(G, \text{SP})$, and we define a set U as

$$\{x \in \mathbb{R}^{n-l+1} \mid x_{n-l+1} \geq 0, |x_p| \leq \frac{x_{n-l+1}}{C}, 1 \leq p \leq (n-l)\}.$$

So we further define a one-one mapping from the describing variables to the cone $c(G, \text{SP})$ $f(x_1, \dots, x_{n-l+1}) : U \rightarrow c(G, \text{SP})$

$$f(x_1, \dots, x_{n-l+1}) = \sum_{p=l+1}^n x_{p-l} e_p + \sum_{p=k+1}^l \frac{x_{n-l+1}}{C} e_p + \sum_{p=1}^k x_{n-l+1} \times e_p.$$

Then we can evaluate

$$\begin{aligned} & \int_{c(G, \text{SP})} e^{-\|x'\|^2} dx' \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_U e^{-\|f(x)\|^2} dx \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_0^\infty \int_{-\frac{x_{n-l+1}}{C}}^{\frac{x_{n-l+1}}{C}} \cdots \int_{-\frac{x_{n-l+1}}{C}}^{\frac{x_{n-l+1}}{C}} \\ & \quad e^{-x_1^2 - \cdots - x_{n-l}^2 - (k + \frac{l-k}{C^2}) x_{n-l+1}^2} dx_1 \cdots dx_{n-l+1} \\ &= \sqrt{k + \frac{l-k}{C^2}} \int_0^\infty e^{-(k + \frac{l-k}{C^2}) x^2} \times \left(\int_{-\frac{x}{C}}^{\frac{x}{C}} e^{-y^2} dy \right)^{n-l} dx \\ &= 2^{n-l} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{x}{C\sqrt{k + \frac{l-k}{C^2}}}} e^{-y^2} dy \right)^{n-l} dx, \end{aligned}$$

where $\sqrt{k + \frac{l-k}{C^2}}$ is due to the change of integral variables. We obtain the conclusion of this lemma by combining this integral result with (7.64). \square

7.13.3 Proof of Lemma 7.7

Proof. Consider any fixed $\delta > 0$. First, we consider the internal angle exponent ψ_{int} , where we define $\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}}$. Then for this fixed δ ,

$$\frac{1-\gamma'}{\gamma'} \geq \frac{\frac{C^2-1}{C^2}\rho\delta + \frac{\delta}{C^2}}{\rho\delta} - 1$$

uniformly over $\nu \in [\delta, 1]$.

Now if we take ρ small enough, $\frac{C^2-1}{C^2} \frac{\rho\delta + \frac{\delta}{C^2}}{\rho\delta}$ can be arbitrarily large. By the asymptotic expression (7.34), this leads to large enough internal decay exponent ψ_{int} . At the same time, the external angle exponent ψ_{ext} is lower-bounded by zero and the combinatorial exponent is upper-bounded by some finite number. Then if ρ is small enough, we will get the net exponent ψ_{net} to be negative uniformly over the range $\nu \in [\delta, 1]$. \square

7.13.4 Proof of Lemma 7.8

Proof. Suppose instead that $\rho_N(\delta, C) > \frac{1}{C+1}$. Then for every vector w from the null space of the measurement matrix A , any $\rho_N(\delta, C)$ fraction of the n components in w take no more than $\frac{1}{C+1}$ fraction of $\|w\|_1$. But this cannot be true if we consider the $\rho_N(\delta, C)$ fraction of w with the largest magnitudes.

Now we only need to prove the lower bound for $\rho_N(\delta, C)$; in fact, we argue that

$$\rho_N(\delta, C) \geq \frac{\rho_N(\delta, C=1)}{C^2}.$$

We know from Lemma 7.7 that $\rho_N(\delta, C) > 0$ for any $C \geq 1$. Denote $\psi_{net}(C)$, $\psi_{com}(\nu; \rho, \delta, C)$, $\psi_{int}(\nu; \rho, \delta, C)$, and $\psi_{ext}(\nu; \rho, \delta, C)$ as the respective exponents for a certain C . Because $\rho_N(\delta, C=1) > 0$, for any $\rho = \rho_N(\delta, C=1) - \epsilon$, where $\epsilon > 0$ is an arbitrarily small number, the net exponent $\psi_{net}(C=1)$ is negative uniformly over $\nu \in [\delta, 1]$.

By examining the formula (7.20) for the external angle $\gamma(G, SP)$, where G is a $(l-1)$ -dimensional face of the skewed cross-polytope SP , we have $\gamma(G, SP)$ is a decreasing function in both k and C for a fixed l . So $\gamma(G, SP)$ is upper-bounded by

$$\frac{2^{n-l}}{\sqrt{\pi}^{n-l+1}} \int_0^\infty e^{-x^2} \left(\int_0^{\frac{\pi}{\sqrt{l}}} e^{-y^2} dy \right)^{n-l} dx, \quad (7.65)$$

namely the expression for the external angle when $C=1$. Then for any $C > 1$ and any k , $\psi_{ext}(\nu; \rho, \delta, C)$ is lower-bounded by $\psi_{ext}(\nu; \rho, \delta, C=1)$.

Now let us check $\psi_{int}(\nu; \rho, \delta, C)$ by using the formula (7.26). With

$$\gamma' = \frac{\rho\delta}{\frac{C^2-1}{C^2}\rho\delta + \frac{\nu}{C^2}},$$

we have

$$\frac{1-\gamma'}{\gamma'} = -\frac{1}{C^2} + \frac{\nu}{C^2\rho\delta}. \quad (7.66)$$

Then for any fixed $\delta > 0$, if we take $\rho = \frac{\rho_N(\delta, C=1)-\epsilon}{C^2}$, where ϵ is an arbitrarily small positive number, then for any $\nu \geq \delta$, $\frac{1-\gamma'}{\gamma'}$ is an increasing function in C . So, following easily from its definition, $\xi_{\gamma'}(y_{\gamma'})$ is an increasing function in C . This further implies that $\psi_{int}(\nu; \rho, \delta)$ is an increasing function in C if we take $\rho = \frac{\rho_N(\delta, C=1)-\epsilon}{C^2}$, for any $\nu \geq \delta$.

Also, for any fixed ν and δ , it is not hard to show that $\psi_{com}(\nu; \rho, \delta, C)$ is a decreasing function in C if $\rho = \frac{\rho_N(\delta, C=1)}{C^2}$. This is because in (7.16),

$$\binom{n}{k} \binom{n-k}{l-k} = \binom{n}{l} \binom{l}{k}.$$

Thus for any $C > 1$, if $\rho = \frac{\rho_N(\delta, C=1) - \epsilon}{C^2}$, the net exponent $\psi_{net}(C)$ is also negative uniformly over $\nu \in [\delta, 1]$. Since the parameter ϵ can be arbitrarily small, our claim and Lemma 7.8 then follow. \square

Acknowledgement

This work was supported in part by the National Science Foundation under grant no. CCF-0729203, by the David and Lucille Packard Foundation, and by Caltech's Lee Center for Advanced Networking.

References

- [1] F. Affentranger and R. Schneider. Random projections of regular simplices. *Discrete Comput Geom*, 7(3):219–226, 1992.
- [2] R. Baraniuk, M. Davenport, R. DeVore, and M. Wakin. A simple proof of the restricted isometry property for random matrices. *Construc Approx*, 28(3):253–263, 2008.
- [3] U. Betke and M. Henk. Intrinsic volumes and lattice points of crosspolytopes. *Monat für Math*, 115(1-2):27–33, 1993.
- [4] J. Blanchard, C. Cartis, and J. Tanner. Compressed sensing: How sharp is the restricted isometry property? 2009. www.maths.ed.ac.uk/tanner/.
- [5] W. M. Boothby. *An Introduction to Differential Manifolds and Riemannian Geometry*. Springer-Verlag, 1986. 2nd edn. San Diego, CA: Academic Press.
- [6] K. Böröczky and M. Henk. Random projections of regular polytopes. *Arch Math (Basel)*, 73(6):465–473, 1999.
- [7] E. J. Candès. Compressive sampling. In *Int Congr Math. Vol. III*: 1433–1452. Eur. Math. Soc., Zürich, 2006.
- [8] E. J. Candès and P. Randall. Highly robust error correction by convex programming. *IEEE Trans Inform Theory*, 54:2829–2840, 2008.
- [9] E. J. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Trans Inform Theory*, 52(2):489–509, 2006.
- [10] E. J. Candès, J. Romberg, and T. Tao. Stable signal recovery from incomplete and inaccurate measurements. *Commun Pure Appl Math*, 59:1208–1223, 2006.
- [11] E. J. Candès and T. Tao. Decoding by linear programming. *IEEE Trans Inform Theory*, 51(12):4203–4215, 2005.
- [12] E. J. Candès, M. P. Wakin, and S. P. Boyd. Enhancing sparsity by reweighted ℓ_1 minimization. *J Fourier Anal Appl*, 14(5):877–905, 2008.
- [13] J. F. Claerbout and F. Muir. Robust modeling with erratic data. *Geophysics*, 38(5):826–844, 1973.

- [14] A. Cohen, W. Dahmen, and R. DeVore. Compressed sensing and best k -term approximation. *J Am Math Soc*, 22:211–231, 2008.
- [15] D. Donoho. High-dimensional centrally symmetric polytopes with neighborliness proportional to dimension. *Discrete Comput Geom*, 35(4):617–652, 2006.
- [16] D. Donoho, A. Maleki, and A. Montanari. The noise-sensitivity phase transition in compressed sensing. *arXiv:1004.1218*, 2010.
- [17] D. Donoho, A. Maleki, and A. Montanari. Message passing algorithms for compressed sensing. In *Proc Nat Acad Sci (PNAS)*, November 2009.
- [18] D. Donoho and J. Tanner. Observed universality of phase transitions in high-dimensional geometry. *Phil Trans Roy Soc A*, 367:4273–4293, 2009.
- [19] D. L. Donoho. Compressed sensing. *IEEE Trans Inform Theory*, 52(4): 1289–1306, 2006.
- [20] D. L. Donoho and J. Tanner. Neighborliness of randomly projected simplices in high dimensions. *Proc Nat Acad Sci USA*, 102(27):9452–9457, 2005.
- [21] C. Dwork, F. McSherry, and K. Talwar. The price of privacy and the limits of lp decoding. *Proc 39th Ann ACM Symp Theory of Comput (STOC)*, 2007.
- [22] A. Feuer and A. Nemirovski. On sparse representation in pairs of bases. *IEEE Trans Information Theory*, 49(6):1579–1581, 2003.
- [23] A. Garnaev and E. Gluskin. The widths of a Euclidean ball. *Dokl Acad Nauk USSR*, 1048–1052, 1984.
- [24] Y. Gordon. On Milman’s inequality and random subspaces which escape through a mesh in R^n . *Geome Asp Funct Anal*, 84–106, 1987.
- [25] B. Grünbaum. Grassmann angles of convex polytopes. *Acta Math*, 121:293–302, 1968.
- [26] B. Grünbaum. Convex polytopes, *Graduate Texts in Mathematics*, vol 221. Springer-Verlag, New York, 2nd edn., 2003. Prepared and with a preface by Volker Kaibel, Victor Klee, and Günter M. Ziegler.
- [27] B. Kashin. The widths of certain finite dimensional sets and classes of smooth functions. *Izvestia*, 41:334–351, 1977.
- [28] B. S. Kashin and V. N. Temlyakov. A remark on compressed sensing. *Math Notes*, 82(5):748–755, November 2007.
- [29] M. A. Khajehnejad, W. Xu, A. S. Avestimehr, and B. Hassibi. Weighted ℓ_1 minimization for sparse recovery with prior information. In *Proc Int Symp Inform Theory*, 2009.
- [30] N. Linial and I. Novik. How neighborly can a centrally symmetric polytope be? *Discrete Comput Geom*, 36(6):273–281, 2006.
- [31] P. McMullen. Non-linear angle-sum relations for polyhedral cones and polytopes. *Math Proc Camb Phil Soc*, 78(2):247–261, 1975.
- [32] M. L. Mehta. *Random Matrices*. Amsterdam: Academic Press, 2004.
- [33] M. Rudelson and R. Vershynin. Geometric approach to error correcting codes and reconstruction of signals. *Int Math Res Notices*, 64:4019–4041, 2005.
- [34] M. Rudelson and R. Vershynin. On sparse reconstruction from Fourier and Gaussian measurements. *Commun Pure and Appl Math*, 61, 2007.
- [35] L. A. Santaló. Geometría integral en espacios de curvatura constante. *Rep. Argentina Publ. Com. Nac. Energí Atómica, Ser. Mat 1, No.1*, 1952.
- [36] M. Stojnic. Various thresholds for ℓ_1 -optimization in compressed sensing. 2009. Preprint available at <http://arxiv.org/abs/0907.3666>.
- [37] M. Stojnic, F. Parvaresh, and B. Hassibi. On the reconstruction of block-sparse signals with an optimal number of measurements. *IEEE Trans Signal Proc*, 57(8):3075–3085, 2009.

- [38] M. Stojnic, W. Xu, and B. Hassibi. Compressed sensing – probabilistic analysis of a null-space characterization. *Proc IEEE Int Conf Acoust, Speech, Signal Proc (ICASSP)*, 2008.
- [39] M. Stojnic, W. Xu, and B. Hassibi. Compressed sensing of approximately sparse signals. *IEEE Int Symp Inform Theory*, 2008.
- [40] S. Vavasis. Derivation of compressive sensing theorems for the spherical section property. *University of Waterloo, CO 769 lecture notes*, 2009.
- [41] A. M. Vershik and P. V. Sporyshev. Asymptotic behavior of the number of faces of random polyhedra and the neighborliness problem. *Sel Math Soviet*, 11(2):181–201, 1992.
- [42] W. Xu and B. Hassibi. Compressed sensing over the Grassmann manifold: A unified geometric framework. Accepted to *IEEE Trans Inform Theory*.
- [43] W. Xu and B. Hassibi. Compressed sensing over the Grassmann manifold: A unified analytical framework. *Proc 46th Ann Allerton Conf Commun, Control Comput*, 2008.
- [44] W. Xu, A. Khajehnejad, S. Avestimehr, and B. Hassibi. Breaking through the thresholds: an analysis for iterative reweighted ℓ_1 minimization via the Grassmann angle framework. *Proc Int Conf Acoust, Speech, Signal Proc (ICASSP)*, 2010.
- [45] Y. Zhang. When is missing data recoverable? 2006. Available online at www.caam.rice.edu/~zhang/reports/index.html.
- [46] Y. Zhang. Theory of compressive sensing via ℓ_1 -minimization: a non-RIP analysis and extensions. 2008. Available online at www.caam.rice.edu/~zhang/reports/index.html.